1-1-2011

Molecular details of the catalytic activity of carboxylesterases

Xiaozhen Yu
University of Mississippi

MOLECULAR DETAILS OF THE CATALYTIC ACTIVITY OF CARBOXYLESTERASES

A Dissertation
presented in partial fulfillment of requirements
for the degree of
Doctor of Philosophy
in the Department of Chemistry and Biochemistry
The University of Mississippi

by

XIAOZHEN YU

May 2011

ABSTRACT

Carboxylesterases (CEs; EC 3.1.1.1) are ubiquitous enzymes responsible for the detoxification of xenobiotics. CEs hydrolyze carboxyl esters into their corresponding alcohol and carboxylic acid. Because of their biological functions, especially their roles in converting inactive prodrugs, such as the anti-cancer drug CPT-11, to their active metabolites, a good understanding of the mechanism of the hydrolysis reaction will give us a better direction for drug design. In this study, we used a multidisciplinary approach (computational simulation, molecular biology techniques and enzyme kinetic methods) to study the dynamic motions of CEs and the potential role of these motions in the catalytic mechanism of CEs. We used a *Bacillus subtilis* protein (pnbCE) as a model and demonstrated that the two loops in pnbCE, coil_5 and coil_21, were important in substrate conversion. Furthermore, we found that a C-C bond side chain rotation of Glu310 was a possible mechanism that the enzyme alternates between its active and inactive conformation. The results of these studies give us new insights about the structure-function relationship of CEs and therefore provide valuable information for approaches toward, for example, CE engineering for selected substrates or design of CE inhibitors.

DEDICATION

This work is dedicated to my parents, without whose tireless encouragement I would have given up long ago.

LIST OF ABBREVIATIONS

ABBREVIATIONS	NAME
AChE	acetylcholinesterase
CE	carboxylesterase
CPT-11/Irinotecan	7-ethyl-10-[4-(1-piperidino)-1-piperidino]carbonyloxycamptothecin
DCCM	dynamic cross-correlation map
hCE1	human carboxylesterase 1/human liver carboxylesterase
hiCE	human intestine carboxylesterase
mAChE	mouse acetylcholinesterase
MD	molecular dynamics
MMTK	molecular modeling toolkit
NMA	normal mode analysis
o-NPA	ortho-nitrophenyl acetate
PDB	protein data bank
pnbCE	*p*-nitrobenzyl esterase
rCE	rabbit carboxylesterase
RMSD	root-mean-square deviation
RMSF	root-mean-square fluctuation
RTB	rotation-translation-block
SN-38	7-Ethyl-10-hydroxy-camptothecin

ACKNOWLEDGMENTS

Faculty, friends, and family members have helped me to complete this dissertation. I would like to express my gratitude to these individuals for their support and assistance.

Dr. Randy Wadkins has been a strong and supportive mentor to me throughout my graduate school career. He demonstrated his faith in my ability to do my job, but also gave me great freedom to pursue my own ideas. His enthusiasm and dedication for science has been a constant reminder of why I started studying science in the first place. He has always taken time to help me with all the difficulties in my research project; has always shown faith in my work; has always been a strong advocate for me. He contributed greatly to this dissertation and I sincerely thank him for all his guidance, encouragement and patience in the past five years.

I am deeply indebted to my committee members: Dr. Wei-Yin Chen, Dr. Michael Mossing, Dr. Susan Pedigo and Dr. Gregory Tschumper. I would like to thank Dr. Chen for his kind support throughout the years. I thank Dr. Mossing for constructive discussions and valuable advices about my molecular biology experiments in my committee meetings. I would like to express my appreciation to Dr. Pedigo for the access to the circular dichroism instrument and her tremendous help and advice for both my career and personal life. I wish to thank Dr. Tschumper, who taught me thermodynamics and gave me insightful discussion about my research project. I also would like to thank our collaborator Dr. Potter in St. Jude children's research hospital for his support and help with my research project. I am grateful to the rest of the faculty members in the Department of Chemistry and Biochemistry, who taught me a lot about how to conduct myself both inside and outside of the classroom.

I would like to thank Zhiyan Sheng, who has been a tremendous supporter. I want to thank members of the Wadkins' group: Yogini Bhavsar, Samantha Reilly, Shana Stoddard and previous member Xu Zhang, for their helpful discussions. I also thank all my friends for their support during the past five years.

Finally, I devote all my thankfulness to my family; my parents, to whom this dissertation is dedicated, have been a constant source of love, concern, support and strength all these years. Making them proud is one of the reasons why I kept myself going for so long. I especially want to thank Dr. Yang's family: Xing, Tring and their lovely kids Tingting and Yan. I thank them for inspiring, motivating and loving me like one of their family members. Without the support of my family, I may have never finished this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

1.1 OVERVIEW

Carboxylesterases (CEs; EC 3.1.1.1) are ubiquitous enzymes that hydrolyze carboxyl esters into their corresponding alcohol and carboxylic acid. Because of the biological functions of CEs, especially their roles in metabolizing various substrates including the anti-cancer drug CPT-11, a good understanding of the mechanism of the hydrolysis reaction will give us a better direction for drug design.

In this study, we used a multidisciplinary approach (computational simulation, molecular biology techniques and enzyme kinetic methods) to study the dynamic motions of CEs and the potential role of these motions in the catalytic mechanism of CEs. The results of these studies give us new insights about the structure-function relationship of CEs and therefore provide valuable information for approaches toward, for example, CE engineering for selected substrates or design of CE inhibitors.

1.2 CARBOXYLESTERASE

Carboxylesterases, as their name implies, are members of the esterase family and also the α/β fold hydrolases superfamily. They have been indentified in various organisms ranging from bacteria to mammals, including humans. They are responsible for the detoxification of xenobiotics (*1*). CEs metabolize a variety of substrate including esterified drugs such as cocaine (*2*), heroin (*2*), Ceftin (*3*) and also the anticancer agent CPT-11 (Irinotecan) (*4*). CEs also catalyze the hydrolysis

of other compounds, such as short- and long-chain acylglycerols and long-chain acyl-CoA esters (*5-7*).

So far, only three CEs have been identified in humans. Among these three, the crystal structure of only one CE isozyme, human liver CE (hCE1), is available (*8*) (*9*). The structures of mammalian CEs are highly homologues with those from other organisms such as bacterial CEs. For example, despite that the sequence of hCE1 has only 33% identity with the bacterial *B.subtilis* CE (pnbCE) (*10*), as shown in Figure 1-1, there is only a 2.7 Å RMSD variation across more than 400 residues of their crystal structures (*11*). The pnbCE exhibits both close structural homology and similar catalytic activity with mammalian CEs (*12*). Also, pnbCE and rCE have comparable activities toward the hydrolysis of o-nitrophenyl acetate (o-NPA) and CPT-11 and both enzymes showed reduced catalytic activity when corresponding amino acid residues were mutated (*12*). Because pnbCE is a bacterial model, it can be easily modified, expressed and purified by using *E. coli* expression systems. Therefore, pnbCE makes an excellent model system for studying the structure-function relationship of human CEs.

As shown in Figure 1-1, these enzymes are structurally similar and share the α/β fold core. Their structures can be divided into three domains: the catalytic domain, the α/β domain and the regulatory domain. There are two disulfide bonds (e.g., Cys87-Cys116 and Cys274-Cys285 in hCE1) present in mammalian CEs, which are also conserved among many other esterases in the family and probably play an important role in maintaining the active conformation of these enzymes.

Of the two mammalian CEs that have crystal structures, hCE1 and a rabbit CE (rCE), both show the active site residues are buried at the bottom of a ~22 Å gorge from the surface of the enzymes(*8-9, 13*). The active site gorge is surrounded by aromatic amino acids. It is believed that

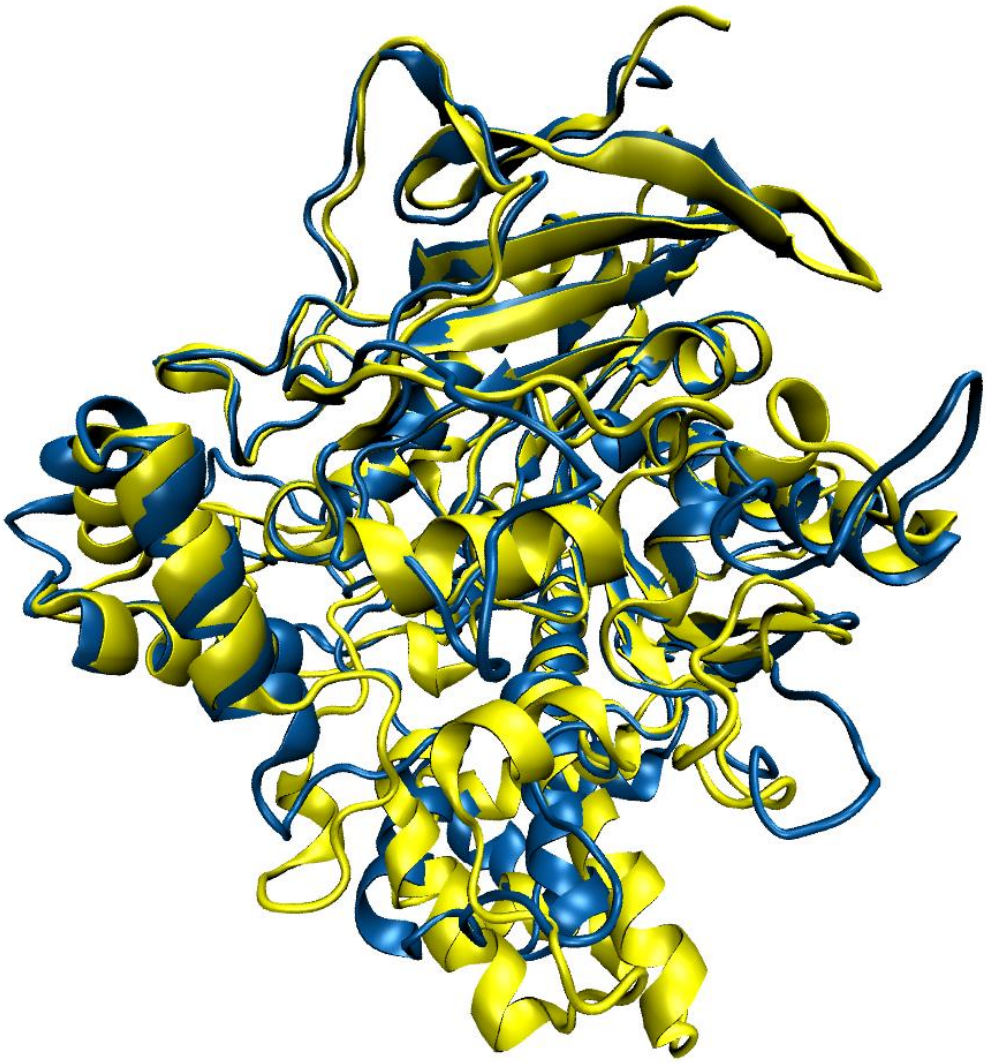


Figure 1-1. The alignment of x-ray crystal structures of pnbCE (PDB code: 1QE3) and hCE1 (1MX5)
The blue cartoon representation structure is pnbCE and yellow structure is hCE1. They both belong to the α/β fold hydrolases superfamily.

the negatively charged gorge likely creates an electrostatic gradient that would pull substrates towards the catalytic residues (*8-9, 13-14*). Although the active site is buried inside this long gorge, these enzymes are able to metabolize such a variety of substrates, showing there must be considerable flexibility and tolerance associated with these enzymes, allowing them to accommodate and hydrolyze structurally diverse compounds.

Previous studies of mammalian CEs proposed that there is a 'side door' adjacent to the catalytic gorge of these enzymes (*13*) (*12*) (*15*) (*16*). This side door is located at the bottom of the substrate-binding pocket and could potentially act as an alternative opening for small substrates, water molecules, or an exit for products (*13*). Similar alternative routes to the active site are also observed in bovine bile salt-activated lipase and acetylcholinesterase (AChE) (*13, 17-18*). Therefore, exploring the potential role of this side door is also important for the mechanism of the hydrolysis by esterases and lipases.

1.3 CPT-11

CPT-11 (7-ethyl-10-[4-(1-piperidino)-1-piperidino]carbonyloxycamptothecin), also known as Irinotecan, is an anti-cancer drug for the treatment of colorectal cancer. It was approved by the FDA in 1994 as a medication given to cancer patients by infusion through a vein.

CPT-11 is a prodrug, whose structure is shown in Figure 1-2. It has to be hydrolyzed by CEs to its active form, SN-38, a topoisomerase I inhibitor. The inhibition of topoisomerase I by SN-38 causes the inhibition of DNA transcription and replication (*19*) and eventually leads to cell death.

There is a problematic diarrhea associated with the use of CPT-11. In some cases, it may lead to severe dehydration. This severe side-effect is related to two human CE isozymes, human

liver CE (hCE1) and human intestinal CE (hiCE). Because hiCE is 60-fold more efficient in the conversion rate of CPT-11 than hCE1, it is believed that hiCE is primarily the enzyme responsible for the activation of CPT-11 (*20*) (*14*). However, the overactivation of CPT-11 to SN-38 in the small intestine by hiCE causes high drug concentrations to be localized in the small intestine. Hence tissue damage to the organ eventually contributes to the delayed severe diarrhea in patients (*14, 21*).

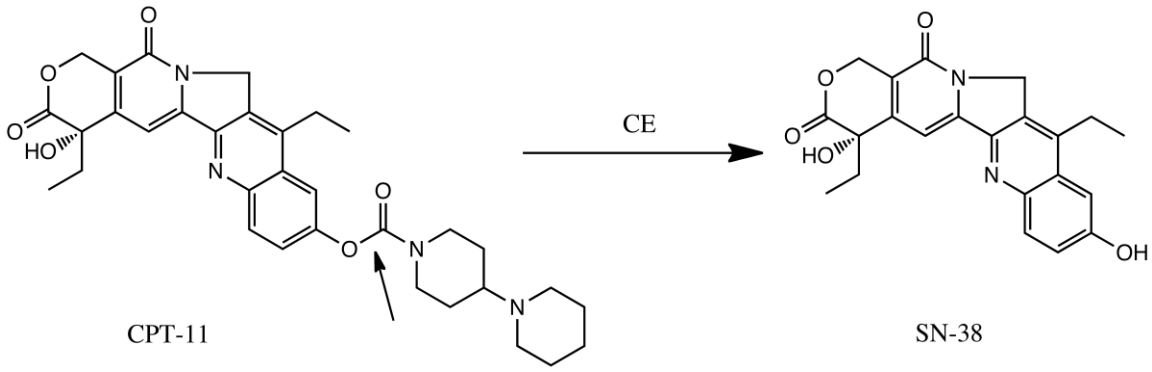


Figure 1-2. The hydrolysis of anti-cancer drug CPT-11 with CE

In order to reduce the side-effect associated with CPT-11, CE inhibitors might be used to specifically inhibit hiCE without affecting the CPT-11 activation by hCE1. Without the activation of the drug by hiCE, hCE1 will be able to activate the drug at a much slower rate, thereby increasing its half-life and reducing its toxic side-effects.

1.4 MECHANISM OF SUBSTRATE HYDROLYSIS

CE cleaves the ester bond in substrates and produces an alcohol and a carboxylic acid, respectively. Structurally, mammalian CEs have four Cys residues and may be involved in forming disulfide bonds (Cys87-Cys116 and Cys274-Cys285, residue numbers are for hCE1).

Among these four, Cys87 is the most conserved residue in CE isozymes. A sequence and structure alignment of a group esterases, lipases and other α/β fold hydrolyses was reported based on crystal structure of the *Torpedo California* acetylcholinesterase and *Geotrichum candidum* lipase (*22*). Based on this structural biology study, Ser221, Glu354 and His468 form a catalytic triad and some neighboring residues may also be incorporated into part of an oxyanion hole, which consists two hydrogen bonds between the negatively charged oxygen anion (called an oxyanion) of the tetrahedral intermediate and N-H group of adjacent Gly residues. These hydrogen bonds are thought to be important in stabilizing the transition state intermediate. This catalytic triad is also confirmed by site-specific mutations analysis. Mutations of Ser221 to Thr221, Glu354 to Ala354 or His468 to Ala468 greatly reduced the CE activity towards substrates. The mutagenesis study confirmed that Glu and His form a putative charge relay system with the other catalytic residue Ser (*23*).

A detailed proposed reaction mechanism is shown in Figure 1-3. One important aspect of the mechanism is the formation of a low barrier hydrogen bond between His and Glu that facilitates the nucleophilic attack by the β-OH group of Ser on the acyl carbonyl group of the substrate (*24*). According to this proposed mechanism, the catalytic triad in the tetrahedral addition intermediate is also stabilized by the low energy barrier hydrogen bond between Glu354 and His468. The overall reaction mechanism of hydrolysis can be divided into following steps.

1. The formation of enzyme-substrate complex. The substrate is positioned into the correct orientation for reaction and the catalytic triad of CE forms the hydrogen bond between Glu and His.
2. The nucleophilic attack is initiated by the oxygen atom of the hydroxyl group of Ser on the carbonyl carbon of the substrate.

3. In this conformation, it is proposed that the negatively charge carboxyl oxygen on the substrate is hydrogen bonded to two nearby Gly and forms an oxyanion hole. Then the original ester bond breaks, and the leaving group picks up a proton from His.
4. The carboxyl portion of the original substrate remains bound to Ser, while the alcohol portion diffuses away.
5. A water molecule comes in as an intermediate. It attacks the acyl-enzyme intermediate to form a second tetrahedral intermediate.
6. Then His donates the proton back to the oxygen atom on Ser, followed by the release of the remaining acid product of the substrate.
7. The acid product is diffused away and the enzyme is ready for the next catalysis cycle.

This mechanism is also known as the "ping-pong" proton transfer hydrolysis mechanism, which is common to all the enzymes so far identified in the esterase family (*11*). The low energy barrier hydrogen bond initiates the whole reaction and is the key step to the stabilization the tetrahedral transition state. In the next step, this tetrahedral intermediate is disrupted by the removal of a proton from His. After the alcohol group diffuses away, the following deacylation step is essentially the reverse of the acylation step.

1.5 USE OF COMPUTATIONAL METHODS IN DRUG DESIGN

As recently reviewed by the Wadkins group, the major classes of known CE inhibitors and *in silico* drug development strategies include using the quantitative structure activity analysis (QSAR) approach (*25*). One traditional method of computationally designing enzyme inhibitors is

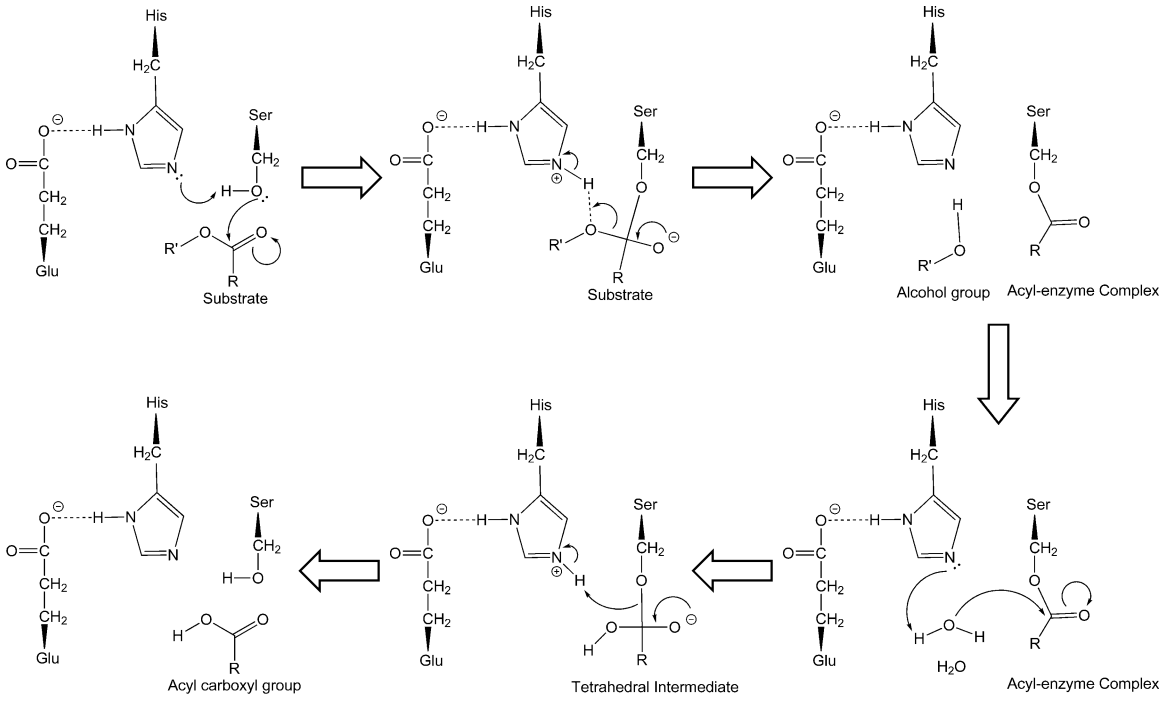


Figure 1-3. Proposed mechanism of substrate hydrolysis by CEs

to dock small organic molecules into the active site of the protein crystal structure in order to calculate the binding interaction between inhibitor and its target. However, this approach is highly dependent on the crystal structures and its use can be limited in practice. First of all, the crystal structure is not always available for an enzyme. Even if it exists, enzymes are usually crystallized with known inhibitors or substrate in the active site, and therefore their structures are usually in compact, closed conformations. Moreover, docking of known inhibitors to a crystal structure of acetylcholinesterase (AChE), an enzyme structurally similar to CE, failed to predict the inhibition constants for known AChE inhibitors (*26*). However, the McCammon group took a more successful approach by incorporating dynamic structures taken from a molecular dynamics simulation instead of using one static crystal structure (*27*). Likewise, the use of dynamic structures becomes very important in resolving the problem for CE as well. Because these enzymes are in constant motion at a temperature of 37 ℃, this approach will allow MD simulation to explore the dynamic conformations of the enzyme and better represent its functional conformations.

In this study, both molecular dynamics simulation and normal mode analysis are used to investigate the possible conformational changes for CEs in order to gain a better understanding of structural information about this family of enzymes. In general, the substrate specificity of CEs is dependent on two structural features: the sizes of their active site gorge and the external opening to the gorge. These important parameters are calculated and measured using molecular dynamics simulations of CEs from rabbit, human liver, human intestine, and a bacterial CE (rCE, hCE1, hiCE and pnbCE, respectively), and will be discussed in detail in chapter 4. In addition, we have also used normal mode analyses to examine their low-frequency motions, which correspond to large conformational changes. The crystal structure of hiCE has not been determined, so a

homology model is generated to be used for this enzyme. It should be noted that earlier homology models of rCEs (*14*) based on the folding of AChE were remarkably similar to the subsequently-determined crystal structure; with a backbone RMSD of ~2 Å. The crystal structures for the other three enzymes (rCE, hCE1 and pnbCE) are available and hence computational analyses can be easily accomplished. However, as we mentioned earlier, like some other enzymes, the crystal structures containing known substrates or inhibitors have an active site that is too small to accommodate other known, larger substrates. For instance, the structure of hCE1 crystallized with a product of benzil hydrolysis becomes too close-packed to allow placement of CPT-11 in the same location. This issue can be resolved by allowing the enzyme structure to fluctuate by using MD simulation. Other regions of structural fluctuation in CEs are the loops that form a putative lid over the entrance to the active site gorge. Normal mode analysis shows these loop motions are consistent among known CE structures. They control the opening and closing of the entrance to CEs active sites, hence probably playing an important role in control of their catalytic activity and substrate selectivity.

The combination of MD and docking has been used for development of specific inhibitors of AChE vs. other serine hydrolase enzymes, and hence we expect this approach has the potential to work for CEs as well. Coupled with 3-dimensional QSAR models as a reduced representation of the inhibitor binding site, these computational methods have led to the prediction of several CE inhibitors (*25*).

The understanding of the dynamic structures of CEs will be crucial for *in silico* docking of CE inhibitors as well as for understanding the origins of substrate specificity in α/β fold hydrolases.

1.6 α/β FOLD HYDROLASES

As we mentioned earlier, CEs belong to the α/β hydrolases fold superfamily. The α/β hydrolases fold was defined by their structural homology. It has been described as consisting of a parallel, eight stranded β sheet surrounded on both sides by α helices. The three dimensional structures of the *fusarium solani* cutinase (FsCUT) and human liver carboxylesterase (hCE1) as well as the structure of acetylcholinesterase (AChE) revealed a very similar topology of the central α-helix/β-sheet fold and the same three residues comprising the catalytic triad: a nucleophile (Ser, Cys or Asp) after β5, an acidic residue after β7 and an highly conserved His residue located after the last β strand (Figure 1-4).

According to the ESTHER database, there are 99 subfamilies in the α/β hydrolases fold superfamily. There are 54 out 99 subfamilies in which at least one protein structure is known. A total of 918 structures have been crystallized so far in the Protein Data Bank (PDB), containing coordinates of native enzymes, or complexes with inhibitors or substrates. These structures represent 256 different proteins in the α/β hydrolases fold superfamily. Although these proteins show little sequence homology (i.e. FsCUT has as little as 197 residues, while mouse acetylcholinesterase (mAChE) has 582 residues (*28*)), the conservation of the topology suggests that these proteins were diverged from a common ancestor (*29*).

One of the reasons that we study CEs, pnbCE in particular, is to discuss implications of our results for the entire α/β fold family of hydrolases. Our results gave us more insight into the possible role of the inserted loops, outside the α/β folding core, in controlling hydrolysis mechanism and substrate selectivity.

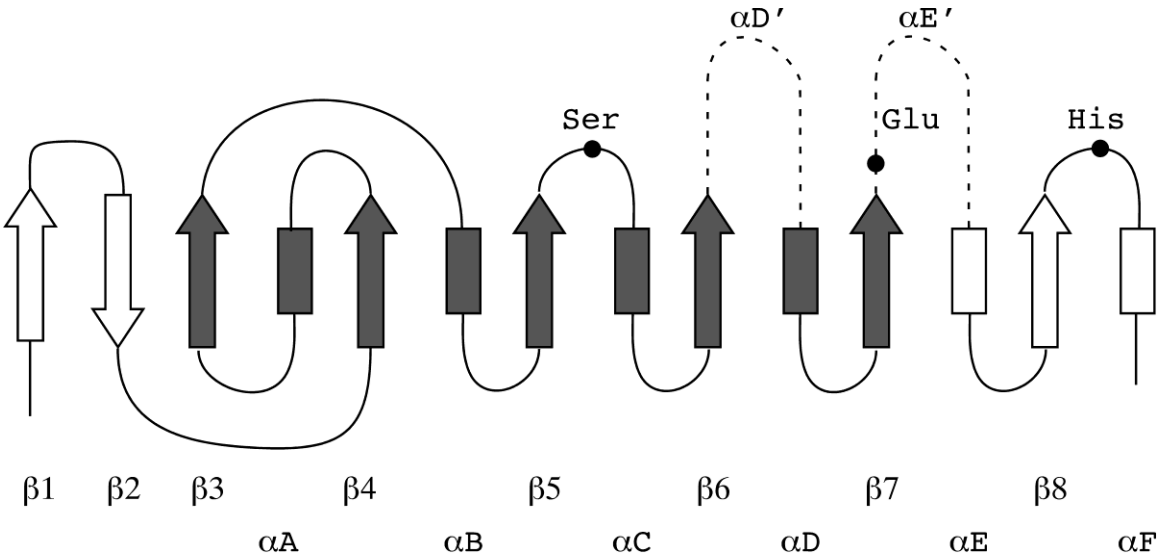


Figure 1-4. Topology diagram of the α/β hydrolases fold.
α-helices and β-sheets are illustrated as rectangles and arrows, respectively. The minimal α/β hydrolases fold is indicated in dark gray. The location of the catalytic triad is indicated in black dots and the corresponding amino acid code. The extra α-helices insertions in front of αD and αE are indicated by dashed lines and labeled as αD' and αE'.

1.7 OTHER POSSIBLE APPLICATIONS OF CES

Our goal is to establish a fundamental understanding of the relationship between enzyme dynamics and substrate activity of CEs. Such an understanding would, for example, allow modulation of CE activity by mutation of known enzymes, resulting in enhanced substrate conversion (*30*). These mutants might be used to protect against exposure to the organophosphate nerve agents Sarin and Tabun that target the related enzyme acetylcholinesterase (AChE). Both Sarin and Tabun irreversibly inhibit AChE and result in the accumulation of acetylcholine in the synapses, leading to uncoordinated nerve activity and death from respiratory failure (*11*). The engineering of CEs that can catalytically inactivate organophosphates could provide effective prevention or early therapeutic response for exposure to these nerve agents. Other engineered CEs might be used for narcotic overdoses. For patients who have overdosed on cocaine, administration of a recombinant CE that could metabolize the drug to inactive metabolites, may be effective in ameliorating the toxicity associated with this agent (*11*). In any event, development of these modified enzymes would be facilitated by a fundamental understanding of the relationship between CE molecular dynamics and substrate activity.

CHAPTER 2 PRINCIPLES, MATERIALS & METHODS

2.1 MD SIMULATION

Molecular dynamics (MD) simulation is becoming one of the principal tools in the study of biological molecules. MD calculates the time-dependent behavior of a molecular system. MD simulation is widely used to obtain information about conformational changes of biomolecules on a time scale that is otherwise inaccessible by most other computational methods, and therefore, it is a useful tool in extending our understanding of biological systems. Along with other complementary computational approaches (i.e., normal mode analysis, to be discussed in next section), MD simulations have become a valuable method for investigating the dynamics of protein structure and function.

The various dynamic processes of proteins have time scales ranging from femtoseconds to seconds. Some of the characteristic time scales of protein motions are shown in Table 2-1 (*31*). Many of these motions play critical roles in biochemical functions (*32*). For enzymatic systems, rapid and localized motions may play a role at the chemical conversion level of enzymatic reactions. Slower and globalized motions are associated with the conformational changes of the enzyme prior to the reaction.

Using MD simulations, we are able to study both thermodynamic and time-dependent (i.e., kinetic) properties. Among the various applications of MD simulation, we are specifically interested in the role of dynamics in enzyme activity (*33*) (*34*). The connection between microscopic simulations and macroscopic properties can be made via statistical mechanics. In

Table 2-1. Time Scales for Protein Motions

Biological event	Amplitude (Å)	Time (s)	MD
Bond-length vibration	0.01 – 0.1	10^{-14} – 10^{-13}	Yes
Rotation of solvent-exposed side chains	5.0 – 10.0	10^{-11} – 10^{-10}	Yes
Hinge bending	1.0 – 5.0	10^{-11} – 10^{-7}	Yes
Rigid-body (helix) motions	10.0 – 50.0	10^{-9} – 10^{-6}	No
Loop motions	10.0 – 50.0	10^{-9} – 10^{-5}	No
Allosteric transitions	1.0 – 5.0	10^{-5} – 1	No

statistical mechanics, one of the most fundamental concepts is the Ergodic hypothesis, which states that the ensemble average of a group equals the time average of an individual component of the group.

The basic idea is that if the system is allowed to evolve in time indefinitely, it will eventually go through all possible states. Therefore, if given enough computational resources, we will eventually be able to generate all the representative conformations of the system. Thus, experimentally relevant information concerning dynamic structures and thermodynamic properties may be calculated using MD simulation. Although in most cases the time scale of a MD simulation might not be able to uncover the entire biological process, nonetheless it allows the characteristics of a model to be probed.

Molecular dynamics simulation is based on solving Newton's second law, F = ma. For a particular particle i,

$$F_i = m_i a_i \qquad \text{(Equation 2-1)}$$

where F_i is the force exerted on the particle i, m_i is its mass and a_i is its acceleration. Meanwhile, the force can also be expressed as the gradient of the potential energy,

$$F_i = -\nabla_i E_p \qquad \text{(Equation 2-2)}$$

Combining these two equations yields,

$$-\frac{dE_p}{dr_i} = m_i \frac{d^2 r_i}{dt^2} \qquad \text{(Equation 2-3)}$$

Equation 2-3 can relate the derivative of the potential energy to the changes in atomic position as a function of time. Choice of an appropriate potential energy function to describe the energy landscape of the biological system becomes vital to the success of a MD simulation.

The general form of potential energy function,

$$E_p = E_{bonds} + E_{\text{angle}} + E_{dihedral} + E_{vanderWaals} + E_{electrostatic} \quad \text{(Equation 2-4)}$$

calculates the potential energy terms as a sum of individual terms describing deviations of bond lengths, bond angles and torsion angles away from their equilibrium values, plus terms that describe van der Waals and electrostatic interactions for non-bonded pairs of atoms. The set of functions used to calculate intermolecular and intramolecular interactions, along with their associated set of parameters, is termed a force field. In this study, we mainly used the Amber force field, which is introduced first here. The basic force field implemented in Amber has the following form (*35*):

$$\begin{aligned} E_p(R) \quad &= \quad \sum_{bond} K_r(r - r_{eq})^2 && \text{Bond} \\ &+ \quad \sum_{\text{angle}} K_\theta(\theta - \theta_{eq})^2 && \text{Angle} \\ &+ \quad \sum_{dihedral} \frac{V_n}{2}(1 + \cos(n\phi - \gamma)) && \text{Dihedral} \\ &+ \quad \sum_{i<j}^{atoms} \frac{A_{ij}}{{R_{ij}}^{12}} - \frac{B_{ij}}{{R_{ij}}^{6}} && \text{Van der Waals} \\ &+ \quad \sum_{i<j}^{atoms} \frac{q_i q_j}{\varepsilon R_{ij}} && \text{Electrostatic} \end{aligned}$$

(Equation 2-5)

where r, θ and φ are bond length, bond angle and dihedral angle values exhibited in the current configuration. The '*eq*' subscript represents the equilibrium, or reference, values for each of those. K_r, K_θ and V_n are bond length, bond angle and dihedral angle constant, respectively. The parameter n is multiplicity and γ is the phase angle for the dihedral angle parameters. R_{ij} is the distance

between atom i and j. A_{ij} and B_{ij} characterize Lennard-Jones potentials. q_i and q_j are the partial atomic charges for atom i and j and ε is the effective dielectric constant. The first three terms represent the energy involving covalently bonded atoms and the last two terms represent the non-bonded energy between all atom pairs that are within the cutoff range.

According to Equation 2-3, integration of the potential energy function leads to atomic trajectories in space and time. The general procedural flow chart is shown in Figure 2-1.

The initial structures are usually x-ray crystal structures whose geometry may not correspond to the actual energy minima used in the force field. Crystal structures usually have bond angles and lengths distortions, causing bad contacts and therefore high energy. Therefore, an energy minimization step is necessary and failure to do so may lead to instabilities when we run dynamics. During the energy minimization, the temperature of the system is set to 0 K and the minimized structure has no kinetic energy associated with it. In order to allow the biological system to explore its potential energy surface, a certain amount of kinetic energy will be given to the system by assigning an initial velocity to each atom based on an algorithm that takes random values from a Boltzmann distribution. At the same time, target temperature of the system will be set to 300 K in order to simulate an actual environment for biological systems. Since the initially assigned velocities will most likely give you a temperature that is off the target temperature, the dynamics algorithm will scale the velocities up (if $T_{actual} < T_{target}$) or scale them down (if $T_{actual} > T_{target}$) to adjust the difference. The temperature will be periodically checked and adjusted until the target temperature is reached and maintained. After equilibration dynamics, a production dynamics is performed. First, a small time interval Δt is set. Then the process of calculating the instant force (F) and acceleration (a), moving the atom and moving the time forward ($t = t + \Delta t$) will be repeated as long as needed. The production dynamics usually lasts at least several

nanoseconds. In the end, large energy profiles and trajectories files are produced. Several AmberTools scripts and in-house python/perl scripts are used in analyzing the results. These are included in the appendix section.

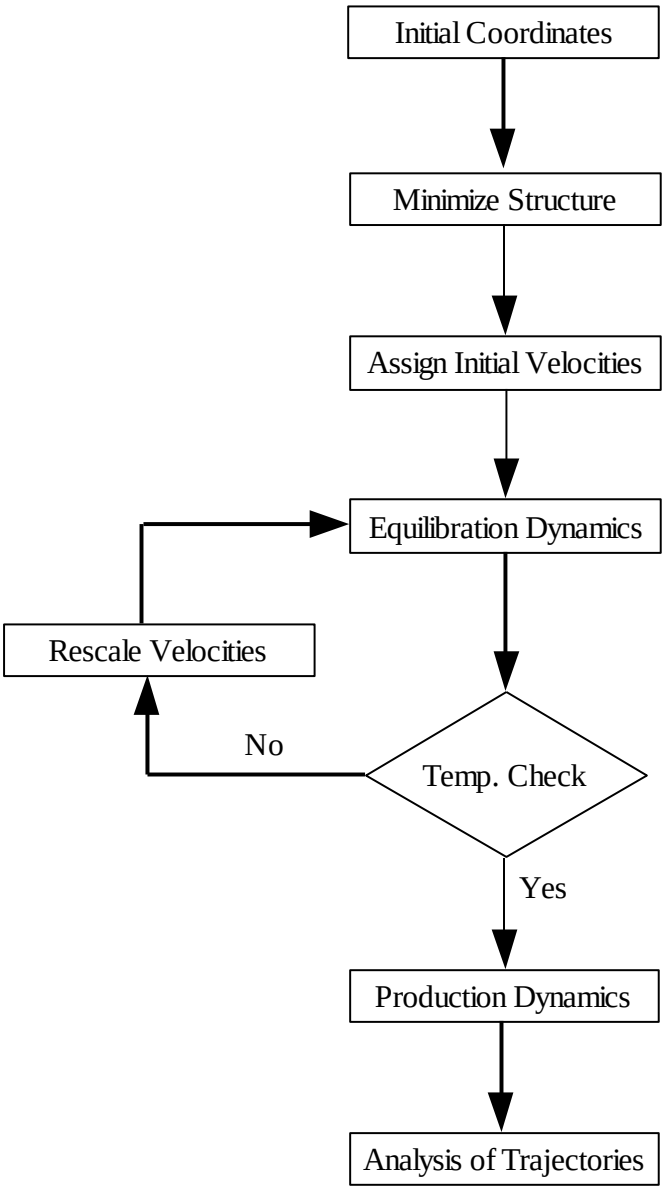


Figure 2-1. Flow chart for MD simulation

2.2 NORMAL MODE ANALYSIS

Although MD simulation is widely used for studying the motions of biomolecules, normal mode analysis (NMA) provides an important alternative to MD simulation because it is much less demanding in terms of CPU time and is able to provide information about low-frequency motions that usually happen on a much longer time scale. NMA of proteins is based on the hypothesis that the vibrational normal modes with the lowest frequencies correspond to the largest movements in a protein that are functionally relevant. It has been shown that some of the lowest-frequency normal modes of some proteins are strongly correlated with their large amplitude conformational changes observed upon ligand binding (*36-38*). The corresponding normal mode coordinates have also been used in studying domain motions of proteins to analyze their molecular dynamics trajectories (*39-41*).

Historically, normal mode analyses of proteins were limited to small proteins due to the use of all-atom description of the biosystem, empirical force fields, and the requirement of the minimization process. These requirements made it computationally difficult with a large system size. However, in recent years, some advanced techniques have been developed and dramatically reduced the computational expenses for NMA (*40*, *42-45*). The most important improvement is the use of anisotropic network model (Hookean potential) (*46*). Along with a reduced representation of protein models (e.g. using Cα atoms to represent each residue of a protein), this model eliminated the energy minimization step, reduced the computational expenses of large protein systems dramatically, and eventually made it possible to set up online NMA servers (*47*)

Normal mode calculation is based on the harmonic approximation of the potential energy surface near a minimum energy conformation. This approximation is one of the most important concepts in molecular mechanics because it allows a simple analytical solution to an oscillating

system. For proteins, their dynamic motions can be seen as a superposition of normal modes, fluctuating around a local minimum energy conformation (e.g. x-ray crystal structure). Most of the amplitude of the atomic displacement is associated with the lowest-frequency normal modes. In general, the solution of the equations can be solved by diagonalizing the Hessian matrix (the mass-weighted second derivatives of the potential energy matrix). The eigenvectors of the matrix are individual normal modes, and the eigenvalues are the squares of their corresponding frequencies. Traditionally, normal mode analysis is performed using detailed empirical force field such as Amber (*48*) or CHARMM (*49*). There are a few factors limiting the size of the system using these traditional all-atom force field approaches. First, since the size of the Hessian matrix that has to be diagonialized is 3Nx3N, where N is the number of atoms in the system. This diagonalization process is time-consuming and has been the limiting step in normal mode analysis. Second, an energy minimization step is required in advance to obtain a structure where the harmonic approximation is valid. Furthermore, typically the minimization is performed in the absence of any solvation models and a gas-phase minimum is usually not a reliable representative of a solution-phase structure as found in most biological systems. However, a new description of the potential energy has been introduced for normal mode analysis that has several advantages allowing for overcoming the limiting factors in all-atom force field approaches. It is a form of a single-parameter Hookean pair-wise potential, which was shown to generate low-frequency normal modes as accurate as those obtained with traditional force fields (*46*),

$$E_p = \sum_{d_{ij}^0 < R_c} c(d_{ij} - d_{ij}^0)^2 \qquad \text{(Equation 2-7)}$$

where d_{ij} is the distance between two atoms *i* and *j*, d_{ij}^0 is the distance between the atoms in the original input structure, *c* is a spring constant of the Hookean potential, which is assumed to be the same among all types of atom-pairs and R_c is an arbitrary cut-off (it is set to 8 Å in most cases),

beyond which pair-wise interactions are not taken into account. This approximation assumes that the reference structure represents the minimum energy conformation, so no energy minimization step is required. Moreover, this Hookean potential eliminates the differences between interacting atom-pairs by using the same spring constant and long-distance interaction are neglected as well by assigning the cut-off distance R_c. Although these approximations tend to break down at high frequencies, they are shown to have little influence on the lowest-frequency normal modes, which corresponds to most biologically relevant conformational changes (*46*) (*50*) (*51*).

The original elastic model (*46*), along with later introduced Gaussian Network Model (GNM) (*52-54*) further reduced the computational cost of normal mode analysis. First, this model used Cα to represent all the residues in the protein and then replaced the interactions between residues in close distance with springs. Furthermore, the introduction of the building block approximation (also known as 'rotation-translation-block', RTB model), groups several neighboring residues into single super-residues. Instead of using the transitional Cartesian coordinates, the rigid-body rotations and translations of the super-residue are used in GNM calculations. Again, for large proteins, the use of super-residue block instead of single residues is expected to have little impact on the lowest-frequency modes (*51*).

The normal mode calculations used in these studies are carried out on the ElNémo web server (*55*). This server provides a web interface that utilizes the Elastic Network Model and the RTB approximation to compute and analyze lowest-frequency normal modes for biomolecules. The principle input structure is in Protein Data Bank (PDB) format. The calculation is done by constructing the interaction matrix for the elastic network model and computing its 100 largest eigenvalues (lowest frequencies) and their corresponding eigenvectors (normal modes). For each individual mode, the degree of collectivity of its movement and the mean square displacement of

all residues is generated.

In ElNémo, the concept of degree of collectivity (scales from 0 to 1) means the fraction of residues that are significantly affected by a given normal mode. For collective movements that are thought to relate to functional conformational change of the protein (*51*), the degree of collectivity tends to be close to one, whereas for localized motions where the normal mode movement only involves few atoms, the degree of collectivity approaches zero. While low-frequency normal modes are expected to have collective characters, because of the nature of the calculation, computed low-frequency normal modes sometimes happen to be localized and therefore have a low degree of collectivity value. These motions, as often observed in crystallographic protein structures for N- or C-terminal of the protein, usually have no significant biological impact and can be ignored. In this study, only the low-frequency normal modes with relatively high degree of collectivity are considered for analysis.

ElNémo also calculates the displacement ΔR value for each residue of the protein in a individual normal mode. B-factors can be calculated from ΔR^2 by using the relationship $B = (8\pi^2/3)* \Delta R^2$ and linearly scaled to the observed B-factors in x-ray crystal structures. The comparison between calculated and observed crystallographic B-factors is a measurement of how well the normal modes are representing the protein's flexibility in its crystal environment.

Normal mode analysis combined with a simplified representation of protein model has shown to be a powerful tool for the study of dynamic properties and conformational changes of biological systems. In this particular study, we used NMA to identify the flexible regions of our protein structures and planned our molecular biological and enzyme kinetic experiments according to these findings. Combined with experimental approaches, NMA helps us get a better understanding of the structure-and-function relationship of our protein system.

2.3 SITE-DIRECTED MUTAGENESIS

We proposed that the flexible loops are incorporated into catalytic mechanism of CEs by using both MD simulation and NMA. In the next step, mutation study of these loops is designed to test this hypothesis.

Site-directed mutagenesis using oligonucleotides was first introduced in 1978 (*56*) and its pioneer, Michael Smith, shared the Nobel Prize in chemistry in 1993 with the inventor of PCR, Kary B. Mullis (*57*). The basic idea is the use of PCR-based mutagenesis protocol to create point-mutation, deletion or insertion on the original DNA template. Because of the exponential growth nature of PCR, after a number of thermal cycles, a significant amount of mutated DNA fragment or plasmid will be generated. After either getting rid of the original DNA template or purifying the PCR product from the reaction mixture using gel electrophoresis, the resulting mutated DNA can be processed with other recombinant molecular biology methods.

In this study, all the deletion and insertion mutants are generated by using the Phusion site-directed mutagenesis kit (FINNZYMES). It is one of the standard PCR-based site-directed mutagenesis protocols; however, it uses uncomplimentary primer pairs as opposed to the complimentary primer pair approach (such as the QuickChange site-directed mutagenesis kit from Stratagene). For plasmid manipulations, the main difference is that QuickChange approach generates a nicked, circular DNA with a linear amplification rate, while Phusion approach generates a linear DNA with an exponential amplification rate.

The target of the site-directed mutation is the pnbCE gene that is cloned into the BamHI restriction site of the pTriEx-3 vector (Novagen). The PCR amplification of target plasmid follows the general procedures of PCR. However, the primers are phosphorylated at the 5’ ends to eliminate the need for a separate phosphorylation step before ligation. The primers designed for

this protocol are uncomplimentary primer pair, so the reaction product is a blunt-end linear DNA fragment. Since plasmid DNA has a much higher transformation efficiency comparing with linear DNA, a ligation step is required in order to circularize the PCR product. A list of primers used in this study is shown in Table 2-2.

1. Incubate the PCR product with DpnI restriction enzyme overnight at 37 ℃ to destroy all the original template plasmid. Then purify the linear PCR product from an agarose electrophoresis gel by cutting the band at the right size.
2. Circularization of the mutated PCR product by using Quick T4 DNA ligase (New England Biolabs).
3. Transformation into *OrigamiB(DE3)pLacI E. coli* cells.
4. DNA plasmid is purified by electrophoresis.
5. Expression and purification of the mutated proteins.

The detailed descriptions of PCR, restriction digestion, ligation, gel electrophoresis and other experimental methods are discussed in the following sections.

Table 2-2. The list of primers used in this study

Name	Sequence
[coil_5]del-F	5’(phosphate)-CGCCAGTCCGAGGATTGCTTG-3’
[coil_5]del-R	5’(phosphate)-AGACGGCTGCGGGCAAATAGAG-3’
[coil_21]del-F	5’(phosphate)-ATTACGGATGAGGTGAAACAGCTTTCTC-3’
[coil_21]del-R	5’(phosphate)-CAATCCGTCCAGATTTCCAAAGAC-3’

2.4 POLYMERASE CHAIN REACTION

The polymerase chain reaction (PCR) is a molecular biological technique to amplify a specific region of DNA across several orders of magnitude, generating a large amount of copies of a particular DNA sequence. Since its introduction in 1985 (*57-58*), the development of PCR method has had a major impact in the way molecular studies are carried out in research laboratories. The PCR method is based on thermal cycling, meaning cycles of repeated heating and cooling of the DNA sample and polymerase replication of the DNA sequence. Both primers and polymerase are the key components of the PCR reaction. As the reaction progresses, the generated DNA product is used as a template for the next reaction cycle, so that the DNA template is exponentially amplified.

The variants of typical PCR can be used to perform a wide array of applications in molecular biology, including DNA sequencing, cloning/sub-cloning of cDNA, generic screening and site-specific mutagenesis.

A typical PCR set up requires several components, including:

- DNA template: A DNA sample that contains the region to be amplified.
- Two primers: short nucleotide oligos that contain sequences complementary to the target region of the DNA template. In this study, we used the Phusion mutagenesis kit that has the uncomplimentary/phosphorylated primer pair design.
- DNA polymerase: Typically, a thermostable polymerase is used. Phusion Hot-start High Fidelity polymerase (Finzymes) is used in this study.
- Deoxynucleoside triphosphates (dNTPs): The building blocks from which the DNA polymerase synthesizes new DNA strands.
- Divalent ions: Mg^{2+} or Mn^{2+} is usually used in PCR-mediated DNA mutagenesis.

- Reaction buffer: Usually reaction buffer is provided to create a stable environment for polymerase and optimum chemical environment for the reaction.

The PCR is usually carried out in a reaction volume of 20~100 μL in PCR tubes in a thermal cycler. Most PCR tubes are specially designed thin-walled reaction tubes that have optimal thermal conductivity to allow for rapid thermal equilibration. Typically 20 to 40 repeated thermal cycles are performed. A detailed PCR procedure is illustrated in Figure 2-2 and described below:

- Initialization step: This reaction is heated to a temperature of 95~98 ℃ and is held for 30 seconds to 3 minutes prior to the thermal cycle.
- Denaturation step: This is the first step in the thermal cycle which consists of heating the reaction to 95~98 ℃ for 10 to 30 seconds. It disrupts the hydrogen bonds between complementary base pairs, causes the melting of DNA template and yields single-stranded DNA.
- Annealing step: Annealing temperature is one of the most important parameters in PCR experiment. After the denaturation, the reaction temperature is lowered to 50~72 ℃ for 20 to 40 seconds. It allows the annealing of the primers to the single-stranded DNA template to form the 5' end of the complementary strand of the template. The exact annealing temperature used in one particular PCR experiment is determined by the melting temperature of the primer pair, which is affected by the number of complementary base-pairs between the template and primer, primer C/G composition, primer concentration, salt concentration and many other factors. There are several equations used for calculating the empirical melting temperature of a DNA duplex. For example, the Nearest-neighbor method

(*59*) is one of the commonly used equations for estimating the melting temperature of a primer,

$$T_m = \frac{\Delta H}{\Delta S + R \ln(C_1 - \frac{C_2}{2})} \qquad \text{(Equation 2-8)}$$

where ΔH and ΔS are the standard enthalpy and entropy, respectively, for the formation of the DNA duplex from two single-stranded DNA. C_1 is the initial concentration of the primer and C_2 is the initial concentration of the original DNA target strand. R is the gas constant (1.987 cal/mol K).

- Extension step: Usually 72 ℃ is used in this step for most polymerases. In the extension step, the DNA polymerase adds dNTPs to the complementary strand in the 5' to 3' direction by condensing the 5'-phosphate group of the dNTPs with the 3'-OH group at the end of extending DNA strand. The extension time is usually determined by the length of the DNA sequence to be amplified. It is also depended on the type of polymerase used. 30 seconds per 1k base-pairs is usually a good estimation for the Phusion Hot-start HF polymerase.
- Final extension: After repeating the previous three steps for 20 to 40 times in a thermal cycle, a final extension step is performed at 72 ℃ for 5 to 15 minutes to ensure all the remaining single-stranded DNA is fully extended.
- Final hold: Hold the temperature at 4 ℃ for indefinite time to allow temporary storage of the reaction.

After the thermal cycle, a DpnI digestion reaction is used to get rid of the original DNA template. DNA samples purified from most *E. coli.* strands are methylated, while PCR generated DNA products are not methylated. Since type IIM restriction endonucleases, such as DpnI, are

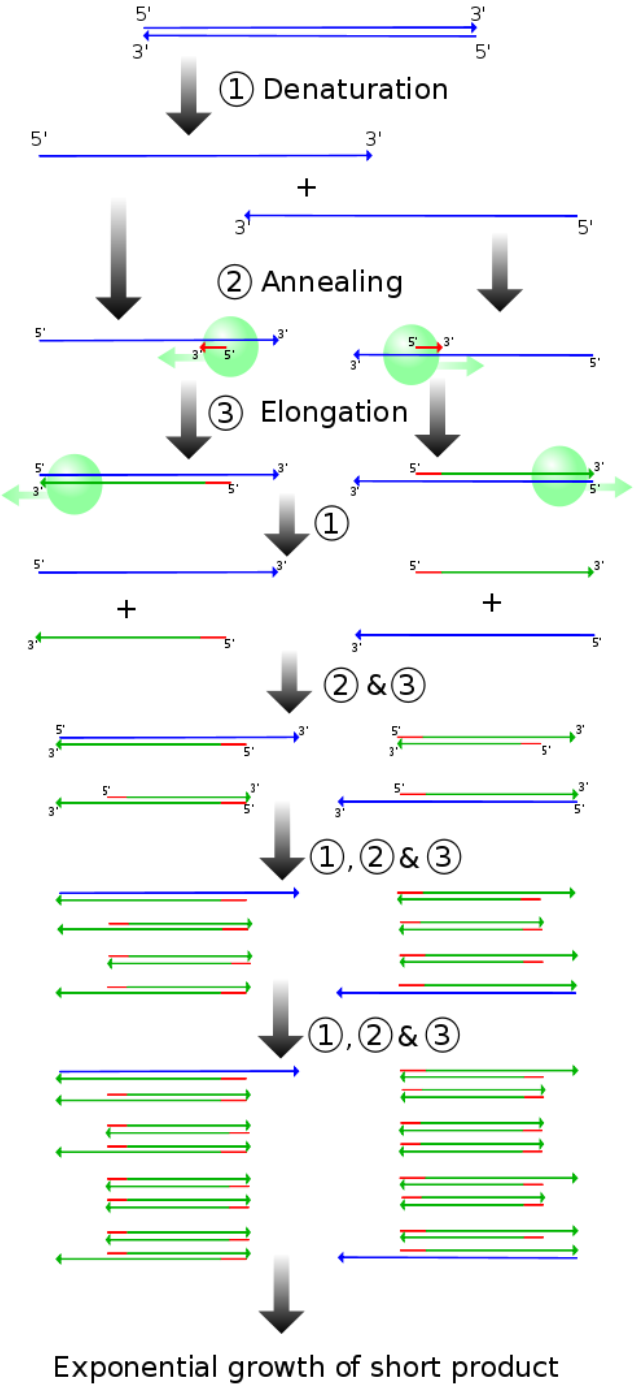


Figure 2-2. Schematic steps of the PCR cycle

able to recognize and cut methylated DNA, they can be used to destroy the original DNA template without affecting the newly synthesized PCR product.

The post-digestion product of PCR experiment can be checked by running an agarose electrophoresis gel along with a DNA ladder. A new band with the target molecular weight is expected from the gel.

2.5 RESTRICTION DIGEST

The discovery of restriction endonucleases is a breakthrough in the development of molecular biology (*60*) (*61*). As part of the natural bacterial machinery, these restriction endonucleases cut double stranded DNA at restricted sites. These enzymes provide us an important tool to study and manipulate DNA by generating DNA fragments consistently at certain sequence sites.

For each restriction enzyme, its recognition sequence produces difference in the length, sequence and strand orientation of digestion products. For example, EcoRI digestion produces "sticky" ends,

G|A A T T C
C T T A A|G

Whereas SmaI digestion produces "blunt" ends,

CCC|GGG
GGG|CCC

There are four classes of restriction endonucleases: types I, II, III and IV. They all recognize specific DNA sequences and differ in their subunit composition, cleavage position and cofactors.

An example of restriction digestion protocol used in this study is shown here,

1. 10x reaction buffer 5 μL

100x BSA	0.5 μL
DNA sample	5 μL
ddH_2O	39.5 μL

2. Add 1 μL of restriction enzyme. Mix thoroughly.
3. Incubate at 37 ℃ for 1 hour.
4. Verify the digestion product by gel electrophoresis and purify the linear product using a gel extraction kit.

A list of restriction endonucleases that are used in this study is shown in Table 2-3.

Table 2-3. The list of restriction endonucleases used in this study

Enzyme	Sequence	Cutting site
BamHI	GGATCC	5'- G ↓ GATCC -3' 3'- CCTAG ↑ G -5'
XhoI	CTCGAG	5'- C ↓ TCGAG -3' 3'- GAGCT ↑ C -5'
DpnI	GATC (methylated)	5'- GA ↓ TC -3' 3'- CT ↑ AG -5'

2.6 GEL ELECTROPHORESIS

Gel electrophoresis is a technique used for separating a mixture of DNA, RNA or protein molecules through a gel matrix in an electrical field. The gel material is a cross-linked polymer allowing for the separation of macromolecules. The composition and porosity of the gel is decided by the size and composition of the macromolecules to be separated and analyzed. Most DNA analyses use agarose as the gel matrix while protein separations usually use different concentrations of acrylamide and a cross-linker as the gel matrix.

By applying an electric field to the molecules in the gel matrix, the electromotive force (EMF) will move the sample molecules based on their charge and mass. For DNA molecules, because they are naturally negatively charged by their sugar-phosphate backbone and their charge to mass ratio is roughly the same, they will migrate in agarose gel from the cathode toward the anode based at different speeds based on their mass.

Unlike DNA, the shape and natural charge of protein molecules varies. Protein samples with the same mass will migrate at different rates in an electric field. Therefore, they have to be denatured with sodium dodecyl sulfate (SDS) to be uniformly negatively charged. Since the binding of SDS to protein ratio is proportional to the protein size (1.4g of SDS per 1g of protein), the rate at which the denatured proteins migrate in the polyacrylamide gel is only related to its size. This gel electrophoresis technique is called sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

After the gel electrophoresis is completed, the separated molecules in the gel can be visualized after a staining process. In this study, ethidum bromide is mostly used for DNA analysis and coomassie brilliant blue is used for protein analysis. Upon binding with DNA or protein molecules, an ethidium bromide molecule fluoresces under ultraviolet light and results a visible band in contrast with the background. A photograph of the gel can be taken under ultraviolet light conditions for qualitative/quantitative analysis.

There is usually a ladder (molecular weight maker) run parallel to the analyte molecules. This ladder contains a mixture of molecules of known sizes. Because the rate of migration is proportional to the size of the molecule, the ladder can be used to compare with the unknown molecules to estimate their molecular weight.

2.7 LIGATION

The process of joining linear DNA fragments together through covalent bonds is called ligation. It is widely used in molecular biology laboratories for recombining fragments of DNA into new DNA molecules. Specifically, DNA ligation creates a phosphodiester bond between the 3'-OH end of one nucleotide and the 5'-phophate end of another nucleotide.

The enzyme used in ligation is T4 DNA ligase, which was originally isolated from the T4 bacteriophage (*62*). With ATP in the reaction, T4 DNA ligase is able to recombine both sticky and blunt ends (see section 2.4 restriction digest). However, higher enzyme concentration and longer reaction time is required for blunt ends ligation.

An example of restriction digestion protocol used in this study is shown here,

1. 10x ligation buffer 5 μL

   DNA sample 10 μL

   ddH_2O 35 μL
2. Add 1 μL of T4 ligase (with a optimal concentration of ATP). Mix thoroughly.
3. Incubate at 16 ℃ for 1 hour.
4. Take ligation product directly to transformation without heat inactivation (heat inactivation dramatically reduce transformation efficiency).

2.8 TRANSFORMATION & EXPRESSION

The ligated circular plasmid DNA needs to be transformed into cells in order to express its corresponding protein. We used the *Origami B(DE3)pLacI* host strain (EMD), which are derived from a *lacZY* mutant of BL21 to enable precise control of expression. The *lacZY* deletion mutant of BL21 enables adjustable levels of protein expression throughout all cells in a cell culture. Its *lacY*

mutation allows uniform entry of IPTG into all cells in the population, which produces a concentration dependent, homogeneous level of induction. By adjusting the IPTG concentration added to a culture, expression can be regulated from low levels up to robust, fully induced levels.

In order to be stably maintained in the cell, a plasmid DNA molecule must contain an origin of replication, which allows it to be replicated in the cell independent of the replication of the cell's own chromosome. Because transformation usually produces a mixture of relatively few transformed cells and an abundance of non-transformed cells a method is needed to identify the cells that have acquired the plasmid. The method usually consists of using a plasmid that contains a gene that gives the bacterial cells resistance to an antibiotic that they are naturally sensitive to. The cells are then plated on media that contains the antibiotic; thus only the transformed cells are able to grow. Cells that did not take up the plasmid are killed in the media.

In this study, the pnbA gene is originally inserted into the pTriEx-3 vector (EMD), which is a vector specially designed to be compatible with *Origami B(DE3)pLacI* expression system. The pTriEx-3 vector is a high copy plasmid and has three antibiotic resistant genes (kanamycin, tetracycline and chloramphenicol).

The detailed transformation & expression protocol is also listed here,

1. Thaw a 20 μL aliquot of *Origami B(DE3)pLacI* cells on ice for 2~5 minute.
2. Add 5 μL ligation product into the cells. Mix thoroughly by gently shake the tube a few times.
3. Incubate on ice for 10 minutes.
4. Incubate the tube at 42 ℃ for exactly 30 seconds.
5. Sit on ice for 2 minutes.

6. Add 80 μL room temperature SOC media to the tube and incubate at 37 ℃ with shaking at 250 rpm for 1 hour.
7. Plate the cell culture on a SOC agar plate with the appropriate antibiotic concentrations (15 μg/ml kanamycin, 12.5 μg/ml tetracycline and 34 μg/ml chloramphenicol, respectively).
8. Incubate 12~16 hours at 37 ℃.
9. Pick ~10 colonies from the agar plate and grow 5 mL cell cultures in SOC media with antibiotics. Incubate at 37 ℃ with shaking at 250 rpm to an OD_{600} of approximately 0.5.
10. Save 1 mL culture from each sample and then induce the reminding 4 mL cultures with IPTG and grow them for another 3 hours.
11. Spin the cultures for 15 minutes at 6k rpm and process the cell pellet with Bugbuster (Novagen).
12. Spin the culture for 20 minutes and collect the supernatant.
13. Perform a CE assay with o-NPA as the substrate for CE screening. The samples with CE activity will proceed to the protein purification step with the saved 1 mL cell cultures.

2.9 CE ASSAY

A spectrophotometric assay was used in determining CE activities. Generally, a constant amount of protein sample was incubated with 3 mM of o-NPA, as the substrate, in 50 mM HEPES buffer (pH 7.4). The changes in absorption per minute were measured spectrophotometrically at

420 nM using a Cary-4 UV-VIS spectrophotometer. The CE activities were expressed as micromoles of o-NP produced per minute per milligram of CE.

Negative control was performed by using the pTriEx-3 vector by itself. After going through the transformation and expression protocol described previously, the negative control shows no detectable CE activity.

2.10 PROTEIN PURIFICATION

We used Isoelectric Focusing (IEF) as the primary protein purification method in our study. It is a technique for separating proteins based on their isoelectric point (pI) value.

When a protein is in a pH region below its pI value, it will be positively charged and so will migrate towards the cathode. However, as it migrates through a gradient of increasing pH, the protein's overall charge will keep decreasing. When the protein reaches the pH region that corresponds to its pI, it has no net charge and stops migrating. As a result, the proteins become focused into sharp stationary bands with each protein positioned at a point in the pH gradient corresponding to its pI. The IEF technique is capable of extremely high resolution with proteins that differ in pI value by as little as 0.01.

Our protein crude extractions are dialyzed overnight in 10 mM HEPES (pH7.4) buffer to get rid off the salt in protein solutions. Then the protein samples are distributed over a pH gradient medium that is created by ampholytes (Bio-Rad). Wild type pnbCE has a pI of ~4.7 and are focused in pH 3~5 ampholytes, while the mutants are focused in different pH ampholytes based on their individual pI values. By applying an electric current to the IEF camber, proteins start migrating to their corresponding pI regions. The whole process takes about 5 hours to finish. After collecting the protein samples into 20 fractions, a CE assay is used to identify the fractions that

contain our enzyme. The right fractions are pooled together and dialyzed in 10 mM HEPES (pH7.4) to get rid of small molecular weight ampholytes. The protein concentrations are determined by Bradford assay and their purities are confirmed with SDS-PAGE gels. The purified protein samples are stored in 4 ℃ for kinetic study.

2.11 BRADFORD PROTEIN ASSAY

The Bradford protein assay is a simple and accurate procedure to determine concentration of solubilized protein (*63*). It involves the addition of protein solution with an acidic dye and measurement at 595 nm with a spectrophotometer. By comparing to a standard curve, the measurement provides an estimation of protein concentration.

The dye reagent used in this study (Bio-Rad) is a mixture of dye, phosphoric acid and methanol. Among these components, Coomassie Brilliant Blue G-250 provides the maximum absorbance spectrum which was historically held to be at 595 nm. The dye primarily binds to basic and aromatic amino acid residues (*64*) and the extinction coefficient of the dye-protein complex solution is constant in a certain concentration range (*65*). Thus, Beer's law can be applied for an accurate quantitative measurement of the protein concentration.

The bovine serum abumin (BSA) is used as the protein standard and the procedures of the assay include:

1. Prepare dye reagent by diluting one part Bio-rad protein dye concentrate with four parts distilled water.
2. Prepare 1mg/mL BSA stock solution and prepare 5 standard solutions with the 1mg/mL BSA stock, covering the linear range of 0.2 to 0.9 mg/mL.

3. Pipet 100 μL of each standard solution, along with the protein sample solution, into a test tube.
4. Add 5 mL of the diluted dye reagent to each tube. Vortex and incubate at room temperature for more than 5 minutes but less than 1 hour.
5. Measure absorbance at 595 nm.
6. Plot the absorbance reading of the 5 standard solutions versus their concentrations. These 5 data points are expected to fall into a linear relationship.
7. Based on this standard curve and the absorbance measurement of protein sample, calculate the concentration of the protein solution.

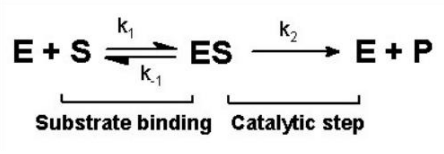
Every sample is assayed in triplicate. All the protein concentrations in this study are measured and calculated using Bradford protein assay.

2.12 ENZYME KINETICS

The study of enzyme catalyzed chemical reactions is called enzyme kinetics. It is usually done by measuring the reaction rates and the effects of varying reaction conditions. In this way, the catalytic mechanism of the enzyme may be revealed. Since we are especially interested in the effects of structural dynamics of pnbCE on its substrate selectivity and catalytic activity, our approach involves changing some flexible features of the enzyme and monitoring their effect on the kinetics. By comparing these data with the wild-type pnbCE, we should get a better idea of the role of these flexible loops of the enzyme in substrate selectivity catalysis.

Previous studies suggested that the hydrolysis of o-nitrophenol acetate by pnbCE exhibits standard Michaelis-Menten kinetics (*12*). This is a simple kinetic model where there is no intermediate, product inhibition or allostericity. In the Michaelis-Menten model, the enzymatic

reaction is assumed to be irreversible, therefore the reaction can be written as:



(Equation 2-9)

With the pseudo-steady-state hypothesis and the assumption that the total enzyme concentration does not change, the initial reaction rate v_0 can be expressed as,

$$v_0 = \frac{v_{\max}[S]}{K_M + [S]} \qquad \text{(Equation 2-10)}$$

where v_0 is the initial reaction rate, Vmax is the maximum rate, [S] is the substrate concentration and K_M is the Michaelis constant. Equation 2-10 can also be converted to,

$$\frac{1}{v_0} = \frac{K_M}{v_{\max}} \cdot \frac{1}{[S]} + \frac{1}{v_{\max}} \qquad \text{(Equation 2-11)}$$

To determine the kinetic parameter K_M and Vmax, a series of kinetic experiments are carried out by varying substrate concentration – [S]. Enzyme concentration is kept at constant and the initial rate v_0 is measured during a short time period. Because [S] and v_0 are the two variables that can be measured in the relationship, Equation 2-11 makes it easier to determine the kinetic constants when data points are used in a Lineweaver-Burk plot. As shown in Figure 2-3, standard Michaelis-Menten kinetics falls into a linear relationship where the slope of the equation is K_M/Vmax and the interception on the y-axis is 1/Vmax.

Although the kinetic parameters can be estimated using the Lineweaver-Burk plot, this method is error-prone and should not be used for accurate calculation in enzyme kinetics. In this study, the kinetic parameters and r^2 values are determined by using non-linear regression with Prism software (Graph Pad) and Equation 2-10.

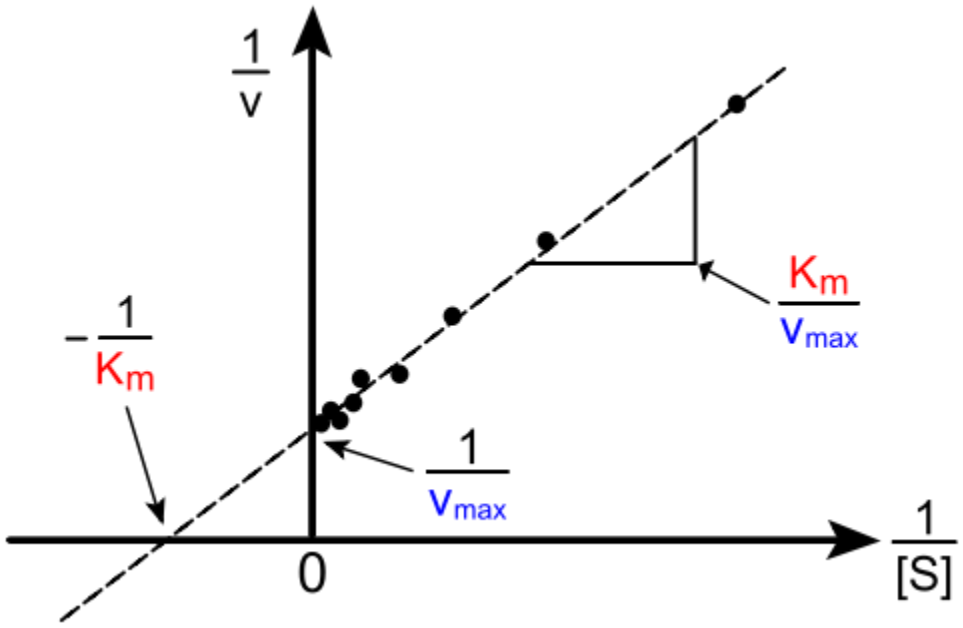


Figure 2-3. The Lineweaver-Burk plot of the Michaelis-Menten kinetics

CHAPTER 3 COMPUTATIONAL DYNAMICS OF PNBCE AND THEIR CONTRIBUTIONS TO SUBSTRATE ACCESS AND CATALYTIC ACTIVITY

3.1 ABSTRACT

Carboxylesterases (CEs) are ubiquitous enzymes responsible for the detoxification of xenobiotics. In humans, therapeutically useful drugs are metabolized by these enzymes, including the anticancer agent CPT-11, the antibiotic Ceftin, and the street drug heroin. Structural dynamics of CEs have critical roles in the mechanism of catalytic drug activation and/or detoxification. In this study, both the low- and high-frequency motions of a CE from *B. subtilis* were examined, with emphasis on what individual domains of the enzyme might contribute to catalytic activity. We used both normal mode calculations and a 10-ns all-atom molecular dynamics simulation to understand catalytic consequences of structural changes within this enzyme. Our results describe low-frequency motions of residues in large loops that may sweep and/or keep substrate in the active site. The higher-frequency motions reveal atomic details of catalytic residue interactions during the simulation. The active site gorge diameter fluctuated from 2.0-7.8 Å, making it capable of accommodating a wide range of substrates. We observe bond rotation between two conformations of Glu310, which plays an important role in formation of a low-barrier hydrogen bond with the active site His399. Such a hydrogen bond would facilitate nucleophilic attack and stabilize the resulting tetrahedral intermediate. In moving between the two conformations, the C-C bond of Glu310 rotates 120 °from an inactive conformation where the hydrogen bond is not made,

to an active conformation. We discuss the implications of our results for the entire α/β fold family of hydrolases.

3.2 INTRODUCTION

Carboxylesterases (CEs) are ubiquitous enzymes responsible for the detoxification of xenobiotics. They are found in organisms ranging from bacteria to man (*23, 66*). In humans, therapeutically useful drugs are metabolized by these enzymes, which impacts upon the efficiency of some drug treatments (*11*). For example, CEs convert inactive prodrugs to active metabolites, a process that is essential for the biological activity of drugs such as the anticancer agents CPT-11 and capecitabine, the antibiotics Ceftin and Vantin, as well as the illicit street drug heroin. However, more commonly, CEs hydrolyze many esterified drugs to inactive products that are then excreted. Agents such as flestolol, meperidine, lidocaine and cocaine, are all hydrolyzed and inactivated by these enzymes. Therefore the efficacy of esterified drugs will be dependent upon the distribution and catalytic activity of different CEs.

Both CE substrates and inhibitors of CEs show selectivity among isoforms from different species, despite the overall structural homology of the enzymes. Examples are given in references (*67-73*), where selected inhibition of 2 human enzymes (hiCE and hCE1) can be accomplished, depending on the structure of the inhibitor. Similar selectivity in the substrates for these enzymes has also been noted (*14*).

The specificity of CEs for a particular substrate or inhibitor depends on the enzyme's molecular structure, and importantly, on the dynamics of conformational substructures present within an enzyme when a substrate is bound (i.e., "conformational sampling" (*74*)). Structural dynamics of CEs, therefore, have critical roles in drug detoxification and can be studied by a

combination of experimental and theoretical techniques. Once a fundamental understanding of the relationship between enzyme dynamics and substrate selectivity is established, modulation of CE activity by mutation of known enzymes may allow for improvements in, and potentiation of, substrate conversion (*30*). Such mutants might, for example, be used as a therapeutic agent for treatment of heroin overdoses or for exposure to nerve agents such as sarin.

Studying the substrate and inhibitor selectivity of a CE from *B. subtilis* has proven it to be an interesting model system for understanding the contribution of enzyme dynamics to the substrate catalysis. The *p*-nitrobenzyl esterase (pnbCE) from this bacterium has a structural homology that is similar to the CE found in rabbit liver (rCE) (*12*). Further, its crystal structure has been determined (*10*), allowing for the 3-dimensional localization of specific residues. Its cloning and expression has made for a simple system to characterize at the molecular level the contribution of specific amino acid residues to both substrate selectivity and the effect of inhibitors. For example, an exploration of the "side door" hypothesis for substrate release by pnbCE has been recently performed, where the kinetics of substrate turnover are shown to be highly dependent on the L362 residue located at the bottom of the active site gorge (*16*).

However, to truly understand the individual residue contributions to the catalytic properties of an enzyme, molecular dynamics (MD) need to be taken into account. The use of MD in the development of enzyme inhibitors for the related α/β hydrolase acetylcholinesterase (AChE) has been recently reviewed (*26*) and these principles also apply to CEs. Briefly, enzymes are in constant motion at temperatures near 37 ℃ and the understanding of the fluctuation is crucial for understanding of the interactions of the enzyme with its substrate or inhibitors. In the α/β fold hydrolase family, the active site residues are at the bottom of a ~22 Å gorge, the walls of which are also fluctuating. This is especially important given that, in the pnbCE crystal structure, the active

site gorge is too small to accommodate large substrates like CPT-11. However, the large CPT-11 molecule is indeed a substrate for pnbCE (*12*). This dilemma can be resolved by structural fluctuations at both the active site opening as well as within the gorge. Taking fluctuations into consideration is particularly useful to understanding how substrates and inhibitors bind to the active site, and such knowledge may, in turn, lead to successful design of enzyme inhibitors (*27*).

Acetylcholinesterase has been the subject of considerable MD simulation for the reasons given above (reviewed in (*26, 75*)). In contrast, only a few studies of the MD of CEs have been reported (*16, 76-77*). Among them a short 400 ps simulation of wild type pnbCE was compared with two thermostable mutants (*76*). These authors noted that the thermostable mutants showed increased populations of low-frequency vibrational modes and suggested that this reflected concerted motions that provided stability to the native state versus the denatured state. However, the connection between these low-frequency motions and catalytic efficiency was not explored. It has become clear in recent years that dynamics have a profound influence on the catalytic properties of enzymes (*74, 78-79*). In this report, we have examined both the low-frequency and high-frequency motions of pnbCE, with emphasis on the question of what individual domains of the enzyme might contribute to the catalytic activity. We have used both a long 10-ns all-atom MD simulation with AMBER 8 (*35*) as well as normal mode calculations (ElNémo; (*55*)) to determine the largest structural changes within the pnbCE. Our results describe both distinct low-frequency and high-frequency motions that involve residues near the active site and the putative "side door" Leu362 that control entry of substrate and exit of hydrolyzed product. We review these motions below and discuss their potential role in the catalytic mechanisms of CEs.

3.3 RESULTS

3.3.1 OVERALL DESCRIPTION OF PNBCE.

The enzyme is an α/β fold hydrolase consisting of 489 amino acids and is folded into three domains: a regulatory domain, an α/β domain and a catalytic domain (*10*). The regulatory domain is comprised of helices αF and the c-terminal helix (see Figure 3-1 for secondary structure designation). The naming scheme in Figure 3-1 is chosen to be consistent within the α/β fold hydrolase family (*29*). Between the regulatory domain and the catalytic domain lies the α/β domain, which is composed of 5 α-helices (αA-αE) and 8 β-sheets (β1-β8) adjacent to each other. The α/β fold which carries the catalytic machinery is the best-conserved structural features in the α/β fold hydrolase family. Besides the α/β fold helices and β-sheets, there are some excursions of varying length occur in each member of the family (primed (') in Figure 3-1 for pnbCE). Within the catalytic domain, three highly conserved active site residues (Ser189, Glu310 and His399) are located at the bottom of a ~22 Å by ~13 Å by ~18 Å cavity, which is similar to all other known CEs (*11*). In the crystal structure of pnbCE, the diameter of the active site gorge is ~3.6 Å wide, which would only accommodate small substrates in specific orientations. From the static crystal structure of pnbCE there is no clear way to understand how a large substrate like CPT-11 could be positioned into the gorge for hydrolysis. Both molecular dynamic simulation and normal mode analysis were used to explore, among other things, the flexibility of the gorge and its impact on substrate selectivity. Similar to the human liver CE (hCE1) structure, the flexibility and chemical environment of the gorge allows pnbCE to act on structurally distinct compounds (*80*)

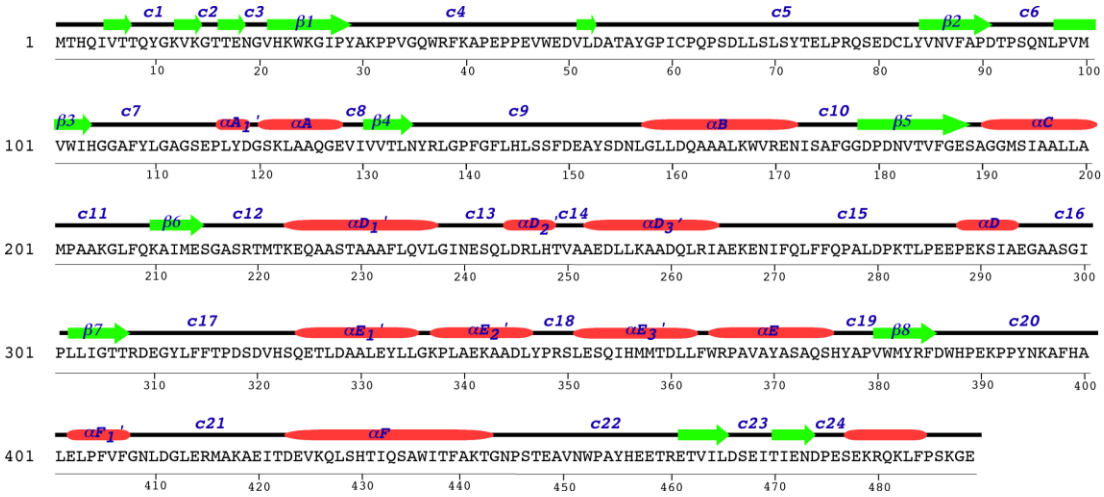


Figure 3-1. The sequence and secondary structural assignment of pnbCE
The secondary structures of pnbCE are illustrated with different color diagrams: green arrows are β-sheets (β1- β8), red bars are helices (αA-αF) and coils/turns in between (c1-c24). The naming scheme is chosen to be consistent within the α/β fold hydrolase family (*29*). The α/β fold helices are labeled as αA, αB, αC, αD, αE and αF. The helices in the excursions are primed (') and labeled as αA_1' before αA; αD_1', αD_2', αD_3' before αD; αE_1', αE_2', αE_3' before αE and αF_1' before αF.

3.3.2 NORMAL MODE CALCULATIONS

The simplest way to attain large-scale, low-frequency motions of molecular structures is through normal mode analysis (NMA) (*81*). These are computationally simple calculations that show global structural changes on a residue-by-residue basis, but do not provide higher-frequency details (these are dealt with in the MD calculations below) (*55, 82-83*).

The first five lowest-frequency vibrational normal modes for pnbCE were generated by the ElNémo web server (*55*). Two of these five modes (mode 7 and 11) with higher degrees of collectivity were used for further analysis since they represented two major global motions. The mean square displacement $<R^2>$ of mode 7 and 11 were calculated against the initial crystal structure. By using the relationship $B = (8\pi^{2/3})<R^2>$ (*84*), predicted B-factors of modes 7 and 11 were calculated, linearly scaled, and plotted against the observed B-factor from the crystal structure (Figure 3-2). Overall these two normal modes exhibited a high correlation with the experimental B-factor. Both modes were able to present a similar overall fluctuation profile and identify the most flexible regions of pnbCE.

It is an interesting dynamic feature of these low-frequency modes that all the active site resides (Ser189, Glu310, His399) are structurally rigid themselves. In contrast, there are several major flexible regions, including coil_17 and coil_20 (Figure 3-2) in close proximity to the active site gorge. These flexible regions constitute the "excursion" regions E' and F' in α/β fold hydrolase nomenclature (*29*), and are common features of large esterases. This flexibility near the active site allows the bottom of the gorge to expand and contract, and likely allows the pnbCE to metabolize a variety of substrates with different molecular sizes and chemical properties.

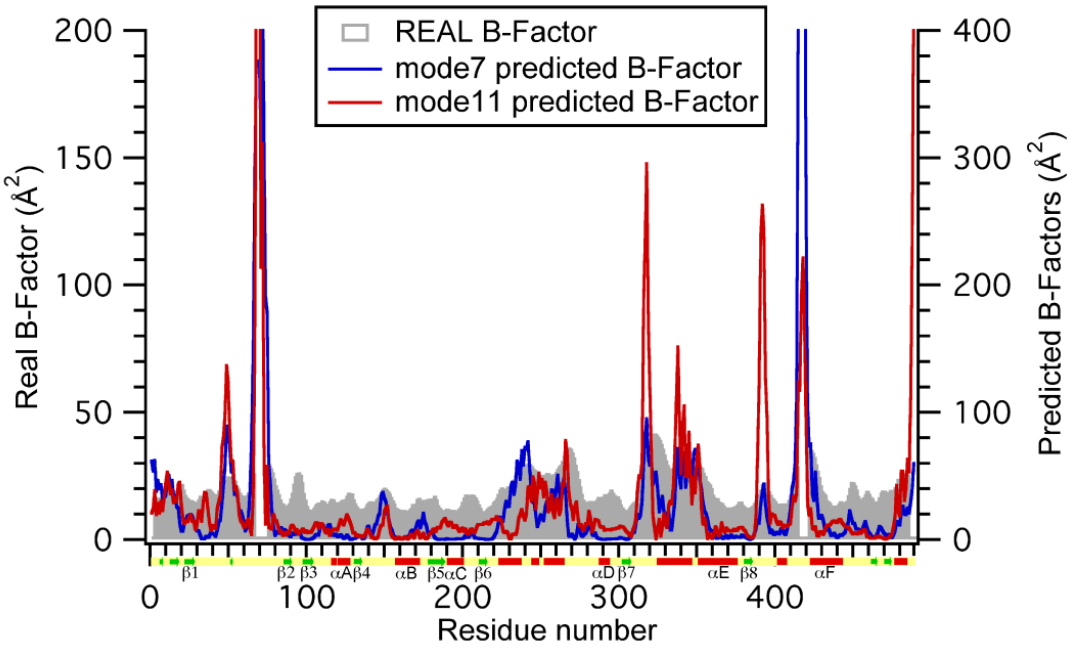


Figure 3-2. The experimental and predicted B-factors of pnbCE
The solid grey area is the experimental B-factor from its crystal structure (PDB entry 1QE3). Red and blue lines are the predicted B-factors from analysis of normal mode 7 and 11, respectively. Predicted B-factors are scaled for comparison purpose. The colored bar under the abscissa is used to represent the secondary structure assignment: green arrows are ß-sheets, red regions are helices and yellow regions are coils and turns. The signature α/β fold hydrolase regions are indicated by αA-αF and β1-β8.

Both modes 7 and 11 represent different characteristic dynamic properties of pnbCE. Mode 7 (Figure 3-3a) was dominated by motions of two loop regions, coil_5 (residues 61-82) and coil_21 (residues 408-422) that are located on the surface of pnbCE. These motions correspond to the poorly resolved regions (residues 66-74 and 414-420) in the crystal structure, indicating their flexibility. The detailed dynamic correlation information will be analyzed later in dynamic cross-correlation map (DCCM) section. As shown in Figure 3-3a, the motions of coil_5 and coil_21 are not only large in magnitude but also highly anti-correlated, while the rest of the protein backbone is essentially rigid. These anti-correlated motions serve to form a flexible cover over the active site, and would necessarily restrict substrate entry. This can be seen more clearly in Figure 4, where the views into the active site, as well as the back side of the active site, during the low-frequency motions of normal mode 7 are shown. The active site residues are shown in red and the "side door" residue Leu362 is shown in yellow in Figure 3-4. The Lee-Richards solvent accessible surface is shown in green. The crystal structure of the pnbCE is shown as the central structure.

The extent of gorge opening and closing at the surface is completely controlled by these motions. Also, in the back view of Figure 3-4, there is the so-called "side door" on the molecular surface, which is directly blocked by Leu362. This "side door" is adjacent to the active sites and is about 3.0 ~ 3.5 Å in diameter, which is similar in size to the narrowest part of the active site gorge diameter. If the "side door" residue unblocks this hole, it could provide a possible entrance for water molecules or exit for small hydrolyzed products. It was indicated in previous studies (*12*, *16*) that this "side door" residue Leu362 plays an essential role during pnbCE-catalyzed hydrolysis of substrates. Although the position of the "side door" residue Leu362 is essentially unchanged during its conformational change in normal mode 7, the dynamic motions of Leu362 that could

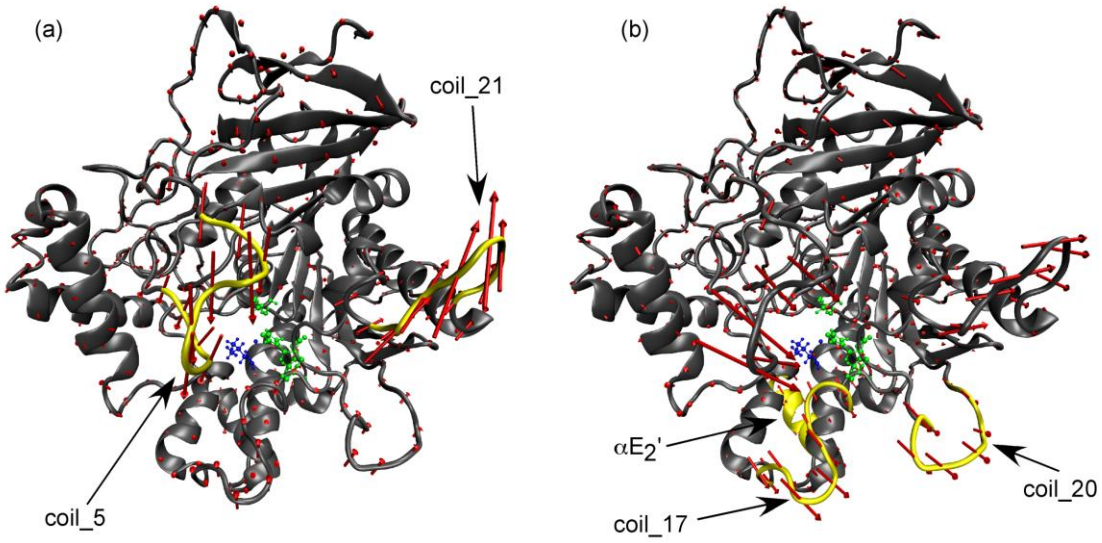


Figure 3-3. Normal mode analysis of pnbCE
The ribbon-and-arrow representation of pnbCE, with red arrows indicating the displacements of the C_α atoms. The grey ribbon (backbone) and CPK (active site) representations are shown. The catalytic triad is shown in green, while the "side door" leucine is in blue. The regions of greatest displacement are labeled and indicated by the yellow section. (**a**) Normal mode 7. The yellow section serves as a "lid" over the active site. (**b**) Normal mode 11. Regions in yellow are close to the catalytic triad and contribute to the size change of the catalytic pocket.

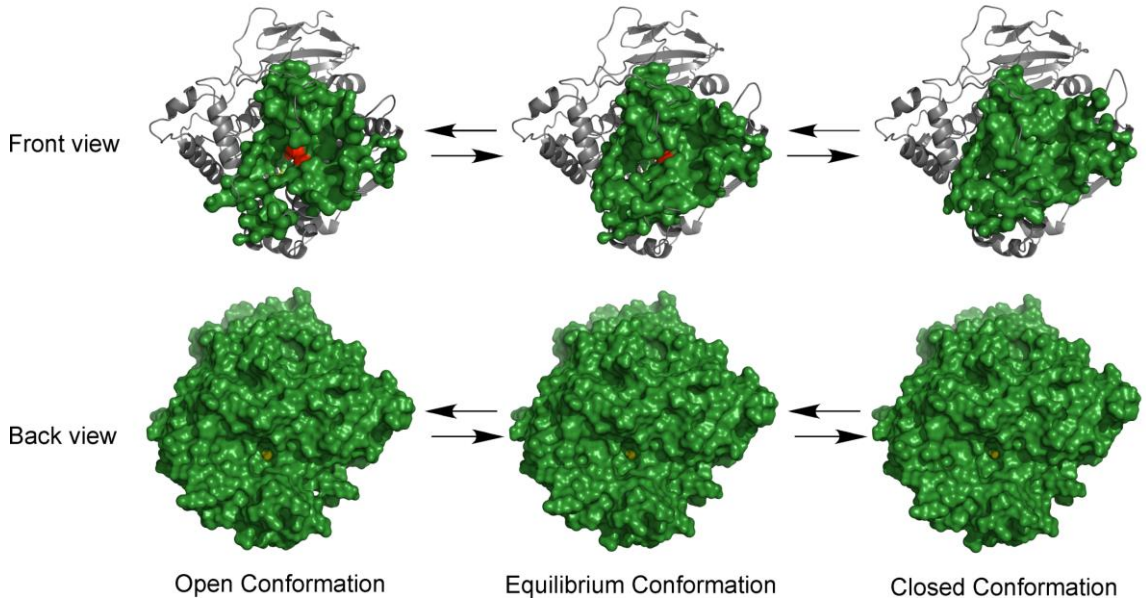


Figure 3-4. The solvent accessible surface of pnbCE
The backbone structure is grey and the active site residues are shown in red. The equilibrium crystal structure, open and closed forms are shown from those calculated using normal mode 7. The opening of the pocket changes dramatically among all three forms and consequently controls substrate entry and exit from the gorge.

possibly facilitate the opening of this “side door”, along with the fluctuations in the active site gorge diameter, is discussed below both in normal mode 11 and in the MD studies.

Normal mode 11 (Figure 3-3b) indicates a second major dynamic motion of pnbCE. The fluctuation amplitudes of three components, αE_2’ (residues 337-346), coil_17 (residues 308-323) and coil_20 (residues 386-401) of pnbCE are enlarged compared to mode 7, indicating a different aspect of dynamic information of pnbCE. In Figure 3-3b, it can be seen that these major differences in motions are near the bottom of the active site gorge. αE_2’ and coil_17 are situated at the back of the active site behind the "side door” residue Leu362. One of the questions around the "side door" concept is how very large substrates such as CPT-11 could exit via this mechanism following hydrolysis to SN-38. Since αE_2’ and coil_17 are located on both sides of Leu362, the anti-correlated motion shown in mode 11 (Figure 3-3b) allows these regions to move away from each other, creating a large space at the bottom of the active site gorge that would allow displacement of Leu362 and create a substantial exit route. This is a potential mechanism for opening the "side door" that would be big enough to release large hydrolysis products. Additionally, as part of the active site gorge, this collective motion of coil_17 may contribute significantly to the fluctuation of the gorge size (see MD simulation below).

3.3.3 NORMAL MODE DYNAMIC CROSS-CORRELATION MAPPING.

The amplitudes of dynamic motions are calculated by the normal mode predicted B-factors. The modes can also be analyzed with respect to direction by dynamic cross-correlation maps (DCCM) to determine which parts of the molecule behave as units. The DCCM make it straightforward to see how the motion of single residue is correlated with all other residues in the enzyme. The DCCM of pnbCE for normal mode 7 is shown in Figure 3-5a while that for mode 11

is shown in Figure 3-5b. A green-red color scale is used to represent the motions, with green indicating correlated motion and red indicating anti-correlated motion. Although, by their very nature, most of the motions calculated by normal mode analysis are correlated to a degree, the DCCM does provide information about the interconnectivity of the secondary structures and their dynamic relationships, which can be seen as off-diagonal crosspeaks.

The DCCM for normal mode 7 (Figure 3-5a) indicates that the predominant correlated segment consists of residues 210-300, which is the region between β-sheet 6 (β6) and α-helix D (αD) of the α/β hydrolase fold. This corresponds to a region containing three α-helices (Figure 3-1) and is the common "excursion" location for α/β fold hydrolases (the region where all members of the family are the least homologous (*29*)). This region is adjacent to the opening of the active site gorge, and is usually the domain that is involved in substrate binding (*28*). While the magnitude of the motion in this region is not as large as the surface loop regions (Figure 3-2), the DCCM maps indicate that this region acts as an intact, isolated region that evolved to maintain its 3D structure, which would be consistent with a role in substrate binding.

One significant off-diagonal segment consists of the anti-correlated motions between coil_5 and coil_21, as the red arrows indicate in Figure 3-3a. Since both loops cover the entrance of the active site gorge, their anti-correlated motions likely plays a role in controlling the process of substrate entering the active site, acting as a lid over the gorge entrance. The other major off-diagonal segment in mode 7 consists of the anti-correlated motions of residues 210-240 and 320-350. As part of the α-helices D and E, they belong to the α/β domain and the regulatory domain, respectively. These α-helices are located at the protein surface around the active site gorge (at the bottom of pnbCE structure in Figure 3-3a) and their anti-correlated motions would greatly contribute to the size fluctuation of active site gorge of pnbCE.

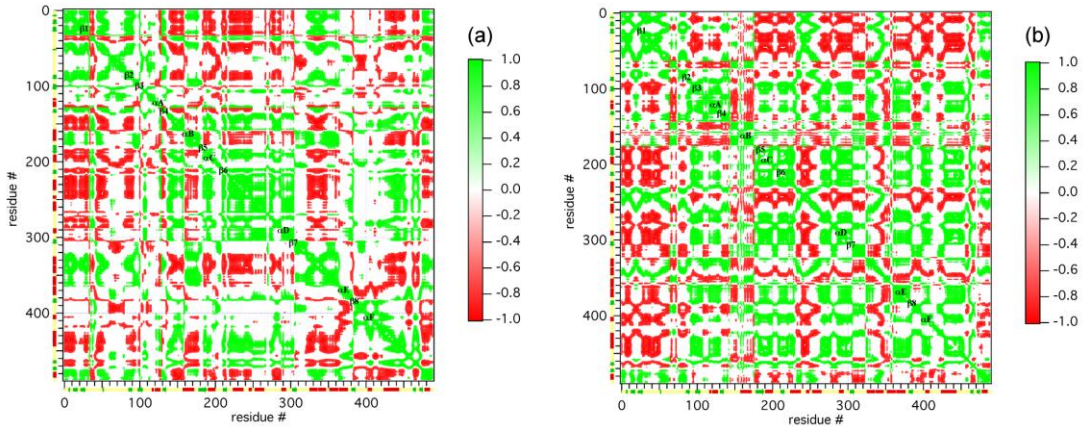


Figure 3-5. Dynamic cross-correlation maps from normal mode analysis
These plots show correlated motions of the C_α atoms between all residue pairs in (**a**) mode 7, and (**b**) mode 11. The green-red scale color bar on the right indicates the extent of the correlation. Green indicates a high level of correlated motions (residue pairs move together in the same direction), while red indicates anti-correlated motions (residue pairs move together in the opposite direction). To highlight the collective motions, only motions in the top 20% of the color scale (-1.0 to -0.8 and 0.8 to 1.0) are shown.

The main difference between mode 7 and 11 in DCCM is that mode 7 (Figure 3-5a) shows more intact and large scale collective motions, while mode 11 (Figure 3-5b) shows more individual secondary structure component motions. This is consistent with mode 11 representing the relatively higher frequency motions comparing with the lowest frequency motions of mode 7. In Figure 3-5b, most of the components in the α/β fold domain move in a correlated fashion, including β3, αA, β4, αB, β5, αC, β6, αD and β7. They are the central components of the α/β folding pattern of pnbCE which carries the catalytic machinery. Thus these correlated motions are responsible for keeping the structural and functional stability of pnbCE.

3.3.4 MOLECULAR DYNAMIC SIMULATIONS.

While NMA allows us to explore the lowest-frequency motions in pnbCE, given the simple empirical energy function used to analyze the motions, no timescale is available for the resulting motions (i.e., the simplified force constants used in NMA are not the true force constants acting on the atoms). Similar loop motions in other enzymes are on the micro- to millisecond time scale are correlated with the catalytic cycle (*78*) and those observed during NMA most likely are on this time scale as well. It was also of interest to examine high-frequency motion on the ps to ns time scale using molecular dynamics (MD) and force constants of greater accuracy and complexity (the AMBER force field). After energy minimization and equilibration of the complete pnbCE molecule, a 15 ns MD simulation was performed. During the simulation process, the pnbCE molecule reached its local minimum energy equilibrium after ~5 ns and fluctuated around -6651.37 ± 16.94 (mean ± sd) kcal/mol (Figure 3-6a), indicating that our model provided a reasonable starting structure for our simulation and was at equilibrium throughout the remaining 10 ns (the time window used for the analysis below).

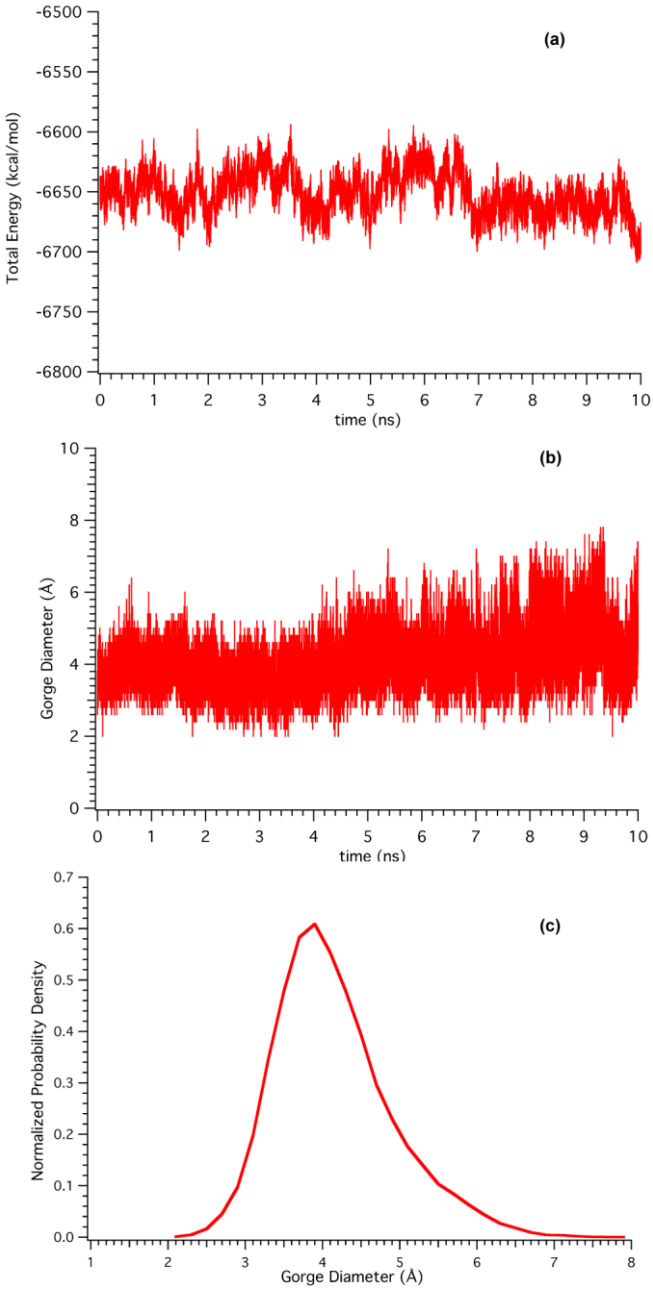


Figure 3-6. MD simulation of pnbCE

(**a**) Total energy fluctuation in kcal/mol during 10-ns MD simulation is presented, indicating equilibrium sampling during the trajectory. (**b**) The pnbCE active site gorge diameter during 10-ns MD simulation. The diameter of active site gorge fluctuates between 2 Å to 7 Å and increases its fluctuation with time during the 10-ns MD simulation. (**c**) The normalized probability density distribution of pnbCE gorge diameter.

Figure 3-6b & c illustrates the gorge diameter fluctuations of pnbCE from our MD simulation, which is a measure of the largest substrate that can reach the catalytic active site triad. Overall, the average gorge diameter fluctuated at 4.05 ± 0.45 Å throughout the simulation. However, the diameter size collapsed to 2.0 Å at its narrowest to as large as 7.8 Å at its widest. The size and fluctuation of the gorge diameter increased during the simulation, from 3.81 ± 0.48 Å (first 2 ns) to 4.71 ± 0.85 Å (last 2 ns). These diameters are significantly larger than that reported for the related acetylcholinesterase (AChE) (*75*), where the gorge diameter fluctuations were found to be 3.04 ± 0.52 Å. Further, we do not see the bimodal switch in gorge diameter that was observed for AChE and that has been postulated to act as a gating mechanism during the hydrolysis of AChE substrates. This is not surprising, inasmuch as CE's are promiscuous enzymes, whereas the AChE is finely tuned for acetylcholine. As we have noted earlier, CPT-11 has a conical molecular volume, with the wider end having a diameter of 7.8 Å and the narrower end a diameter of 3.1 Å (*14*). The gorge fluctuations shown in Figure 3-6b indicate that pnbCE could accommodate a molecule the size of CPT-11 with little difficulty.

3.3.5 ACTIVE SITE AND "SIDE DOOR" RESIDUES.

The catalytic triad, Ser189, Glu310 and His399, is highly conserved among CEs (*85*). As noted in Figure 3-7, it is proposed that hydrolysis is initiated by nucleophilic attack of the β-OH group of Ser189 on the carbonyl group of the substrate (*24*).

According to this theory, the low barrier hydrogen bond between Glu310 and His399 facilitates the initial nucleophilic attack by Ser189 and stabilizes the tetrahedral intermediate as well. Therefore the conformational change involving this Glu-His hydrogen bond is essential for the hydrolysis mechanism. The root mean square fluctuation (RMSF) of the active site and "side

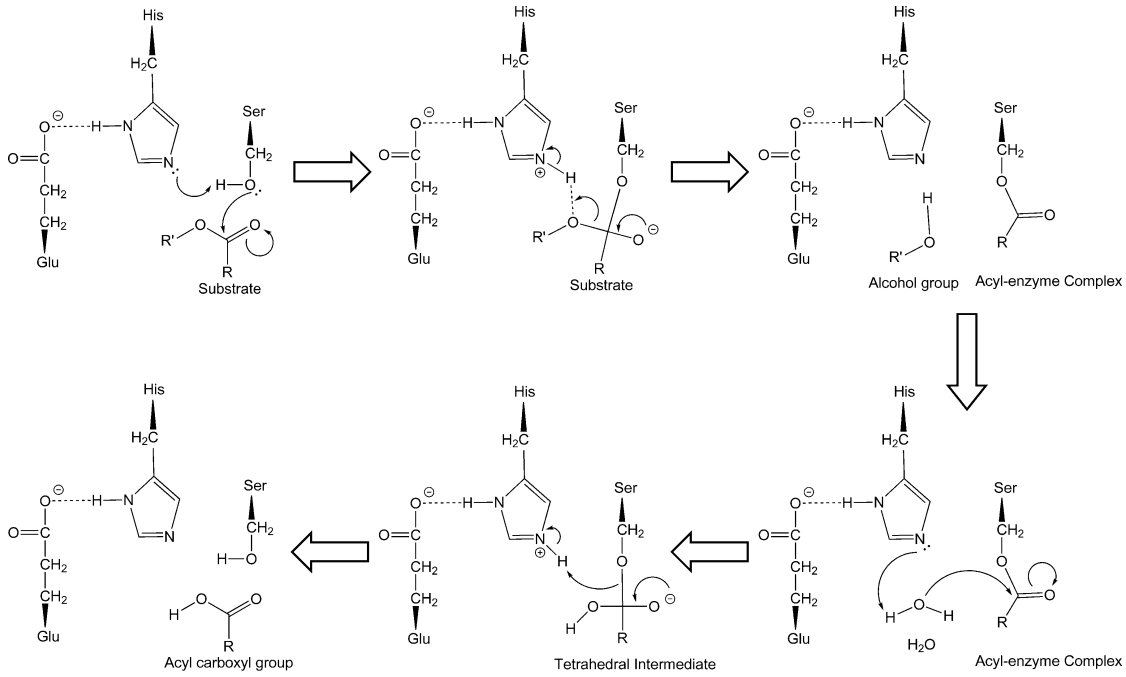


Figure 3-7. Proposed catalytic mechanism of hydrolysis catalyzed by CEs

door" residues from MD simulation revealed interesting features of the active site residue Glu310 and the "side door" residue Leu362. Both residues alternate between two distinctive conformations during the whole of the dynamic simulation. These two conformations are characterized by two distinct RMSF values (vs. the crystal structure); Glu310 fluctuates between positions that are 1.3 Å and 1.5 Å RMSF. Leu362 fluctuates between positions that are 0.7 Å and 2.1 Å RMSF. This is illustrated in Figure 3-8 by focusing on the 1.0 ns region between 6.0 ns to 7.0 ns. In the MD simulation, the conformational details are given in a snapshot at 6.5 ns that shows both Glu310 and Leu362 at their high-RMSF conformation, and another snapshot at 6.9 ns that shows their low-RMSF conformation (Figure 3-9). The two conformations are colored with green and blue to indicate the changes observed.

The large RMSF changes in Figure 3-8 are caused by bond rotation in the two conformations. Glu310 plays an important role in formation of a low-barrier hydrogen bond with the His399 residue. Such a hydrogen bond facilitates nucleophilic attack and stabilizes the tetrahedral intermediate (*24*). In moving between the two conformations, the C-C bond of Glu310 rotates 120 °from an inactive conformation (blue) where the hydrogen bond is not made into an active conformation (green) where it is made.

Also on this time scale, Leu362 rotates 180 °around its C-C bond and adopts two distinct conformations, one of which moves the side chain of the Leu362 away from the active site residue Ser189 by an additional 1.5 Å. This conformational change would make it possible for small hydrolysis products to be released from the enzyme via this "side door". Although speculative, such a rotation as part of the catalytic cycle might also explain the catalytic effects observed in mutation studies of the Leu362 residue (*12*). Mutation of the Leu residue to Lys and Arg would introduce a long, bulky side chain to the residue that would constrain rotational motions and

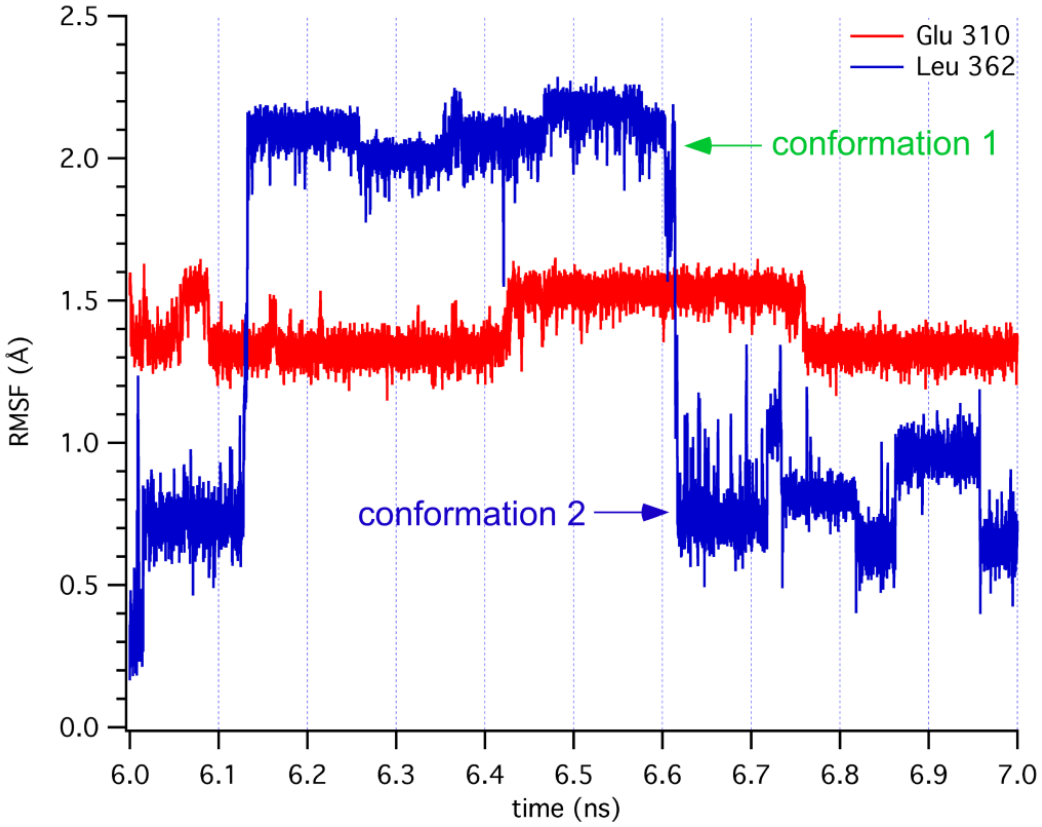


Figure 3-8. The RMSF of two residues Glu310 and Leu362 with respect to their initial coordinates as a function of time for pnbCE

As part of the catalytic triad, Glu310 alternates between two conformations around 1.5 Å RMSF. Meanwhile, the "side door" residue Leu362 also alternates between two major conformations. Both conformation 1 (green) and 2 (blue), as indicated by arrows, correspond to the same two conformations in Figure 3-9.

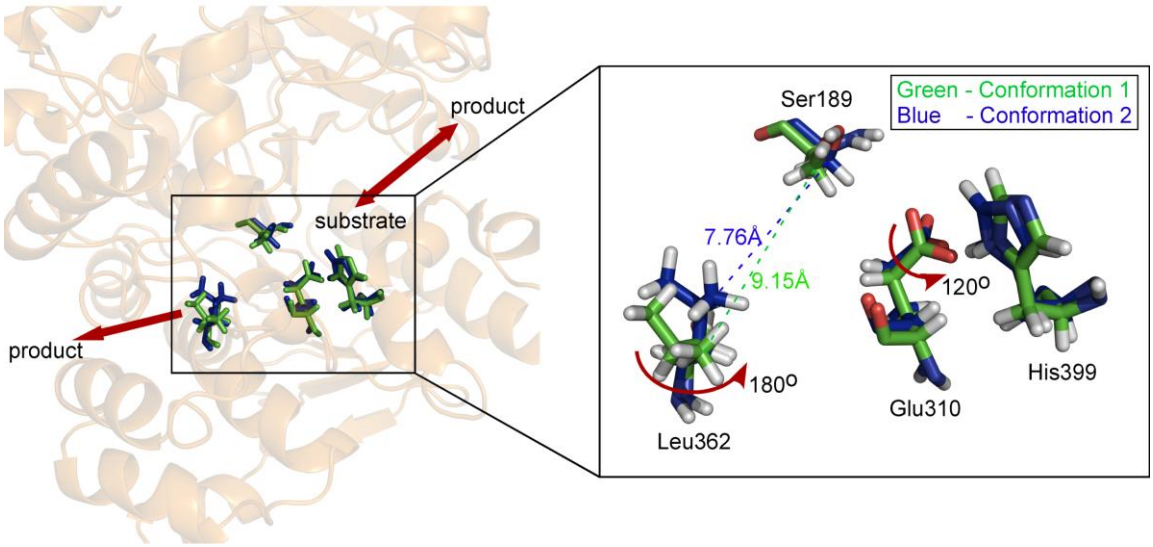


Figure 3-9. A snapshot of pnbCE from the MD simulation showing the active site Glu310 and "side door" residue Leu362 in their two conformations

The transparent cartoon representation in the background is the backbone of pnbCE and red arrows indicate the possible directions of substrate and product movement. The catalytic triad (Ser189, Glu310 and His399) and "side door" (Leu362) from two conformations are highlighted in green (6.5 ns point in Figure 3-8) and blue (6.9 ns point in Figure 3-8). The distances of the closest carbon atoms between Ser189 and Leu362 are shown in dotted lines. Red solid arrows are used to indicate the bond rotations that occur during the conformational change.

potentially block product from exiting this "side door". This would be especially true for large products, which would help to explain why the substitution of Leu with Arg or Lys inhibits CPT-11 metabolism 10- or 5- fold, respectively, but has modest effects on o-NPA catalysis.

3.4 DISCUSSION

In this report, we examined the simulated molecular motions inherent in pnbCE, a carboxylesterase from *B. subtilis*. The global structure of pnbCE is that of a α/β fold hydrolase and we analyzed the dynamic motions of individual components. Our data have important implications for this family of enzymes and their catalytic mechanism.

Recent experimental and computational evidence shows that fast (femtosecond to picosecond) dynamic motions play essential role to enzymatic transition state barrier crossing (*79*). One of the biggest advantages of MD simulation is to explore the fast dynamic motions at the atomic level. The side chain C-C bond rotation of Glu310 provides a possible mechanism for an on/off switch of the low-barrier hydrogen bond with His399 residue, which is essential in the proposed CE hydrolysis mechanism (*24*) (Figure 3-7). Also this bond rotation happens fast enough (200 to 300 ps) to be incorporated into the fast dynamic motions (fs to ps timescale) that links to transition state barrier crossing process (*79*).

Although fast dynamic motions are essential in enzymatic barrier crossing, the overall turnover rate of the enzyme is on the microsecond to millisecond time scale. It is on the same time scale as low-frequency/slow dynamic motions, which are usually corresponding to loop/domain motions. Our normal mode analysis provides a look at the lowest-frequency, largest structural changes in pnbCE, which are generally correlated with catalytic activity. The loops designated coil_5 and coil_21 form a flexible lid over the entry to the active site gorge. These loops are

divergent in the α/β fold hydrolase family and likely help either to control which substrates have access to the active site gorge or which are allowed to exit once the gorge is entered. As a substrate approaches the enzyme, the anti-correlated motion of these loops may serve to orient the substrate for easier access to the active site gorge.

Once inside the gorge, the fluctuating walls act as a diffusion barrier to allow access to the catalytic triad. However, the barrier is much smaller in pnbCE as compared to other, more specialized members of the α/β fold hydrolase since the gorge diameter is much larger. This leads to a greater range of substrates able to arrive at the bottom of the gorge. The enzyme can then release larger products when αE_2', coil_17, and coil_21 move away from the active site in a concerted motion (Figure 3-3b) that would create a large gap in the bottom of the active site gorge. These regions then move back into position for hydrolysis of another substrate.

The catalytic residues Ser189, Glu310 and His399 are located at the bottom of the active site gorge comprised of residues 105–108, 193, 215–216, 268–275, 310–314, 362–363, and 399–400 (*10*). Coil_5 of pnbCE is located right on top of the active site gorge and, as indicated above in normal mode analysis, is one of the most flexible region of the enzyme. It contains the amino acid sequence PQPSDLLSLSYTELPRQSED between Cys^{61}-Cys^{82} residues. This loop is a characteristic structural element that is conserved among the esterase/lipase family (*22*). It is structurally homologous to the large omega loop (Ω loop), Cys^{69}-Cys^{96}, of mouse acetylcholinesterase (mAChE), which was studied extensively by using both computational simulation and experimental measurements (*86-88*). It is believed that the conformational change of the Ω loop of AChE plays a crucial role in ligand binding. Among these Ω loop residues, Tyr72 and Asp74, as part of the peripheral anionic site (PAS) of mAChE, bind cationic substrate as the first step in the catalytic pathway (*14, 89-91*), forming a low-affinity complex at PAS that

accelerates catalytic hydrolysis. It is unknown whether the same binding mechanism exists in carboxylesterases, but given the fact that the coil_5 of pnbCE and the Ω loop of mAChE are homologous (sequence alignment shown below),

```
mAChE     69- CYQYVDTLYPGFEGTEMWNPNRELSEDC -96
                |.|..|.|...:  ||:   .|: ||||
pnbCE     61- CPQPSDLLSLSY--TEL---PRQ-SEDC -82
```

both sequentially and structurally, it is likely that this loop plays a similar role in part of the ligand binding process of CEs.

The conformational changes of the Ω loop of mAChE upon ligand binding have been characterized by a combination of cysteine substitution and site-specific labeling experiments (*87-88*). It was concluded from these studies that Ω loop residues are significantly more mobile than other residues around the active site gorge. Our data is consistent with this idea, and we observe coil_5 dominating the RMSF analysis in both of our normal mode analysis and MD simulation. Furthermore, the structural mobility of mAChE residues L76, E81 and E84 on the Ω loop of mAChE has been studied in these experimental studies (To avoid confusion, mAChE and pnbCE residues are in single letter and three letter formats, respectively). Since both L76 and E84 of mAChE are conserved in pnbCE, their corresponding residues in pnbCE Leu68 and Glu74, along with another polar and negatively charged residue Asp66 which is similar to E81 of mAChE, are selected to calculate their correlation with the active site gorge diameter fluctuation during the 10-ns MD simulation. As shown in Figure 3-10 (a&b), three loop residues (Asp66, Leu68 and Glu74) are on the outer surface of the protein molecule and six gorge surface residues (Tyr109, Met193, Ile270, Glu310, Leu362 and His399) are located around the active site and on the solvent accessible surface of the active site gorge. Among them, Glu310 and His399 are part of the active site triad, Met193 is at the bottom of the gorge, Tyr109 and Ile270 are located on the opposite side

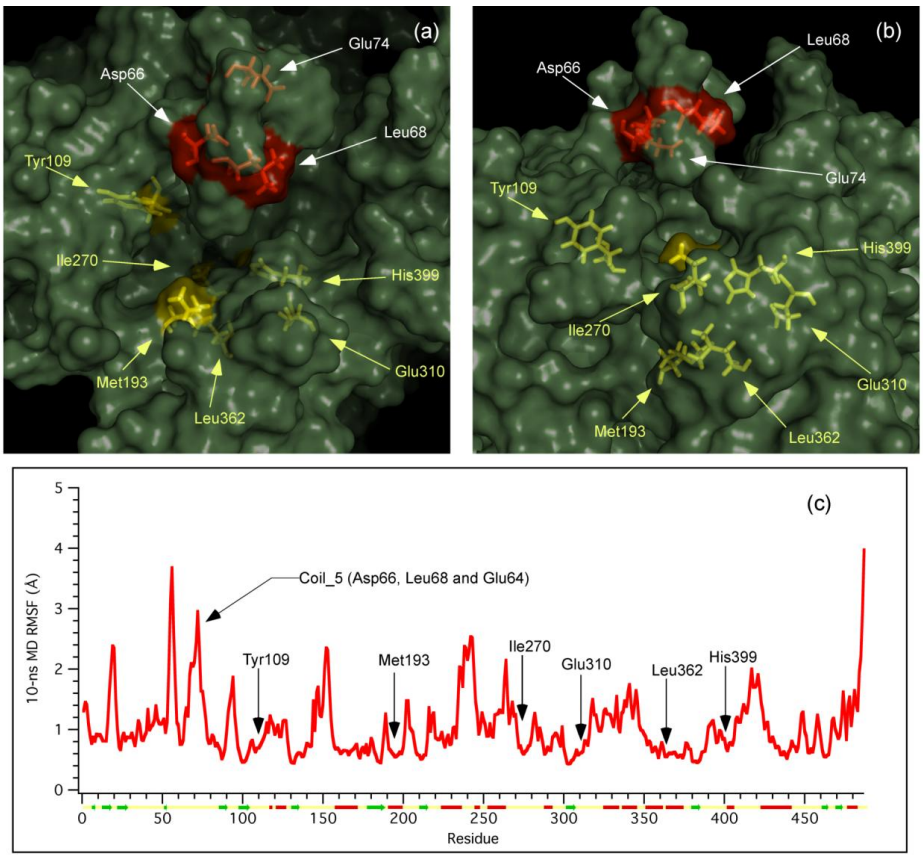


Figure 3-10. RMSD correlation study among selected residues

(**a**) The top view of the active site gorge. The viewers look down the gorge into the active site, with the coil_5 loop region on top of the entrance of the gorge. Residues inside of the active site gorge are highlighted in yellow while loop residues are in red. (**b**) The side view of the same conformation as figure A. (**c**) The RMSD of all the residues after the 10-ns MD simulation. The labeled residues are used in the correlation study.

of the gorge away from the active site, and Leu362 is the "side door" residue. All six gorge surface residues are either at or near the bottom of the RMSD plot during MD simulation (Figure 3-10c), indicating structurally they are comparatively rigid. In contrast, coil_5 is the only region that is not only relatively close to the active site gorge but also possesses one of the highest RMSD values. Hence, the fluctuations of the distances between the three loop residues and the six gorge surface residues (a total of 18 vectors) provide a good measurement of the loop dynamics. A dynamic cross-correlation analysis between each one of the 18 vectors and the gorge diameter was performed. The objective of this analysis is to identify which loop residue motion is correlated with the open/closed motions of the gorge diameter. Table 3-1 shows distances between loop residue Glu74 and the gorge surface residues are highly correlated among themselves (>0.80). Similar correlations of Asp66 and Leu68 can be also seen (tables are not shown). These high correlation values mean that the loop motion involving Asp66, Leu68 and Glu74, is not only a high magnitude motion as the normal mode B-factor (Figure 3-2) and the MD simulation RMSD plot (Figure 3-10c) indicate, but also a highly-organized collective motion.

The first row of Table 3-1 is the correlation between the gorge diameter and the six distance vectors involving Glu74. Along with the same correlation information involving Asp66 and Leu68, the correlation between gorge size and three loop residues is compared in Table 3-2. Among all three loop residues, Glu74 shows a significant correlation (0.3 to 0.5) with the gorge size fluctuation. However, dynamics of Asp66 and Leu68 shows minimal correlation with the gorge size fluctuation despite being on the same loop (Table 3-2). We found that the correlation coefficient between the distance Glu74-His399 and the gorge diameter is 0.5 (Table 3-2), which is similar to the highest correlation coefficient between a C_α-C_α distance and gorge radius in previous MD simulation study of mAChE (*86*). This high correlation means the motion of Glu74

Table 3-1. Normalized dynamic cross-correlation coefficient of 6 distance vectors involving E74 and gorge diameter

$C_{r\rho}$[a]	ρ[b]	Glu74-Tyr109	Glu74-Met193	Glu74-Ile270[c]	Glu74-Glu310	Glu74-Leu362	Glu74-His399
ρ	1.00	0.31	0.39	0.35	0.41	0.32	0.50
Glu74-Tyr109	0.31	1.00	0.80	0.83	0.82	0.81	0.79
Glu74-Met193	0.39	0.80	1.00	0.83	0.91	0.90	0.88
Glu74-Ile270	0.35	0.83	0.83	1.00	0.87	0.83	0.82
Glu74-Glu310	0.41	0.82	0.91	0.87	1.00	0.94	0.96
Glu74-Leu362	0.32	0.81	0.90	0.83	0.94	1.00	0.90
Glu74-His399	0.50	0.79	0.88	0.82	0.96	0.90	1.00

[a] $C_{r\rho}$ - Normalized correlation coefficient (range from -1.0 to 1.0)

[b] ρ - gorge diameter of pnbCE (Å)

[c] Vector distance between two residues

moving away from the active site gorge facilitates the enlargement of the entire gorge, suggesting the open/closed conformational changes of pnbCE are controlled by the loop motion, especially the motion of Glu74 residue.

The result of this analysis is consistent with the experimental observations of mouse Acetylcholinesterase (mAChE) (*87-88*). It shows that the ranking of segmental mobility of the three mAChE Ω loop residues is: E84 > E81 > L76, and also acrylodan modification of substituted cysteine at E84 (corresponding to Glu74 in pnbCE) reduces ligand binding and kinetic parameters over one order of magnitude; however, a similar substitution at L76 and E81 barely alters the kinetics (*87-88*). Our simulation provides a possible explanation to this experimental phenomenon. When a substrate approaches the enzyme, it first binds to Glu74 of the coil_5 loop in pnbCE (E84 of the Ω loop in mAChE). Then this binding process facilitates conformational change around active site gorge, opens up the active site gorge to accommodate the substrate and finally allows the hydrolysis to proceed.

In summary, we combined normal mode analysis and MD simulation to study the dynamic motions of pnbCE and their roles in its catalysis. Our normal mode analysis revealed two distinct low-frequency motions of residues in large loops near the top of the active site gorge that may sweep and/or keep substrate in the active site. A 10-ns MD simulation revealed higher-frequency motions with atomic details as to how the catalytic and loop residues interact during the simulation. Although, the atomic details of motion of the loop region can not be completely explored due to the short time scale of MD simulation, coupled with normal mode analysis, the collective dynamics of this loop region and the active site gorge provide a picture of how this loop facilitates enzymatic conformational change of pnbCE. It provides a direction for additional experimental studies to establish the mechanistic basis of the specific role of each residue on the

loop. Current mutational analysis in our laboratory is ongoing to test the biological significance of these loops in mutants of pnbCE.

Table 3-2. Normalized dynamic cross-correlation coefficient between each of the 18 distance vectors and gorge diameter fluctuation

$C_{r\rho}$	Tyr109	Met193	Ile270	Glu310	Leu362	His399
Asp66	0.01	-0.04	0.20	0.12	-0.09	0.28
Leu68	0.07	-0.11	0.24	0.02	-0.07	0.11
Glu74	0.31	0.39	0.35	0.41	0.32	0.50

3.5 MATERIALS & METHODS

3.5.1 PNBCE STRUCTURE

The structure of *B. subtilis* carboxylesterase (pnbCE) was taken from Protein Data Bank entry 1QE3 (*10*). Two loops, residues 66-74 and 414-420, are missing from the crystal structure and were inserted using Modeller (*92*). The resulting structure was then minimized with respect to energy using the *sander* module of AMBER 8 (*93*). The secondary structure assignment of pnbCE was done by program STRIDE (*94*). Like other carboxylesterses, pnbCE is folded into three domains, a catalytic domain, an α/β domain and a regulatory domain. A total of 15 α-helices, 14 β-sheets, and 25 coils and turns were assigned by the program (see Figure 3-1).

3.5.2 NORMAL MODE ANALYSIS

Normal mode analysis (NMA) was performed on the ElNémo web server (*55*), which is a web interface to the Elastic Network Model that implements the 'rotation-translation-block' (RTB) approximation. The minimized starting structure was used as the input PDB file. The

number of residues that are treated as a rigid body value (NRBL) was determined by the server as a function of protein size to optimize computation speed. A cutoff of 8 Å was used to identify elastic interactions. The first 6 zero-frequency modes contain the overall rotations and translations of the system. Our focus was the 5 lowest non-zero frequency modes (modes 7-11) that were generated. The output of NMA is a series of trajectory files in Protein Data Bank (PDB) format, which include 11 structures by default: the original input structure and ten frames representing the motions between the two conformational extrema of the normal mode.

Modes 8, 9 and 10 have a low degree-of-collectivity value around 0.1, which meant only a few atoms were involved in these motions. Additionally, the motions were all localized within the N- or C-termini of the structure. For the purpose of analyzing structural and functional relationship, these modes were ignored due to the lack of global conformational significance. Meanwhile, modes 7 and 11 had high degree-of-collectivity values of 0.24 and 0.34, respectively. These two modes were used for further analysis.

The mean square displacement $<R^2>$ of all C-α atoms in mode 7 and 11 were calculated from comparison with the input structure. By using the relationship $B = (8\pi^2/3)<R^2>$, predicted B-factors were calculated for both normal modes (*43*). Both $<R^2>$ and predicted B-factors are arbitrary numbers, and can be linearly scaled for comparison purpose. The experimentally observed B-factor was obtained from the x-ray crystal structure of pnbCE. In Figure 3-1, both predicted B-factors were scaled and plotted with the real B-factor, providing a measure of how well the protein's fluctuation in the crystal structure was represented by these two low-frequency normal modes.

Cross-correlation analysis was performed on all the residue pairs in pnbCE for mode 7 and 11, respectively. The cross-correlation coefficient C_{ij}, between atoms *i* and *j*, was computed as

follows:

$$C_{ij} = \frac{< \overrightarrow{R_i} \cdot \overrightarrow{R_j} >}{\sqrt{< \overrightarrow{R_i}^2 > \cdot < \overrightarrow{R_i}^2 >}} \qquad \text{(Equation 3-1)}$$

where R_i and R_j are the displacement vectors for atom *i* and *j*, respectively and the brackets indicate an average over the trajectory. The value of C_{ij} was calculated using an in-house python program. Input of the program is the trajectory file (in PDB format) that was generated by ElNémo server. The result was loaded into a 489 x 489 matrix and the value of C_{ij} ranged from –1 (fully anti-correlated, residue pair moving in the opposite direction) to 1 (fully correlated, residue pair moving in the same direction). In Figure 3-4, only the top 20% of the correlated motions are shown in these dynamic cross-correlation maps (DCCM).

Active site pocket identification and measurements were done on CASTp server (*95*). The weighted Delaunay triangulation and the alpha complex are used for shape measurements (*96*). The calculation used default atomic radii and a solvent probe of radius 1.4 Å. Measurements for both solvent accessible surface (Lee-Richards' surface (*97*)) and molecular surface (Connolly's surface (*98*)) were calculated by the server.

3.5.3 MOLECULAR DYNAMICS SIMULATIONS

MD simulations were performed with the AMBER 8 software suite using the ff94 force field (*48*) and the pairwise generalized Born model (*99*). Energy was minimized for 100 cycles of 2 fs and followed by a 10 ps equilibration cycle, with positional restraints on all atoms. Temperature of the system was gradually increased from 0 K to 300 K during the equilibration process. The MD simulation then used a cutoff of 12 Å for non-bonded interactions and weak-coupling algorithm (*100*) was used to maintain a 300 K constant temperature. A total of 10

ns of MD simulation was performed using a 2-fs time step in conjunction with the SHAKE algorithm (*101*) to constrain bond length between hydrogen and heavy atoms in the system. All the simulations were conducted on the Mississippi Center for Supercomputing Research's Altix cluster computer.

Analysis of the MD data was done on a dual-core Mac desktop. First, the total energy of the 10-ns MD simulation was plotted with respect to time to determine the regime within which the MD reflects statistical equilibrium fluctuations around the energy minimum. Then the *ptraj* program from AMBER was used to calculate the root mean-square fluctuation (RMSF) of 4 residues: active site Ser189, Glu310, His399 and "side door" Leu362, with respect to their starting coordinates.

An in-house python program using the molecular modeling toolkit (MMTK) module (*102*), was used to analyze the MD simulation trajectory. The analysis was performed with a procedure modified from the work of McCammon and colleagues (*103*). As active site Ser189 and His399 are the two residues proposed to make initial contact with the ester bond to be cleaved (Figure 3-7) (*66*), the molecular surface of pnbCE was calculated at each step of the dynamics using a probe radius that was incremented by 0.1 Å until active site Ser189 and/or His399 no longer contribution to the molecular surface. The largest probe radius that could contact both residues was considered to provide the diameter of the active site gorge. The gorge diameter is plotted with respect to time in Figure 3-6b.

The distances between 3 loop residues on Coil_5 and 6 active site gorge residues are measured using the *ptraj* program. For each of the 18 distance vectors, one data point is recorded every 0.1 ps and a total of 100,000 points are recorded during the 10-ns MD simulation. By using equation (1) and a similar cross-correlation analysis Equation 3-2,

$$C_{r\rho} = \frac{< (\vec{r}(t) - < \vec{r} >) \cdot (\rho(t) - < \rho >) >}{\sqrt{< (\vec{r}(t) - < \vec{r} >)^2 > \cdot < (\rho(t) - < \rho >)^2 >}} \qquad \text{(Equation 3-2)}$$

the correlation coefficient $C_{r\rho}$ is calculated for between distance vector pairs (Equation 3-1) and between distance vector and gorge diameter (Equation 3-2). Here r and ρ are the distance vector and gorge diameter, respectively and the brackets indicate an average over the trajectory. The calculated $C_{r\rho}$ is shown in Table 3-1 & 3-2. In Table 3-1, the 7 attributes are gorge diameter and the distances between Glu74 and 6 active site gorge residues (Tyr109, Met193, Ile270, Glu310, Leu362 and His399). The result was loaded into three 7 x 7 matrixes (one matrix for each of the three loop residues - Asp66, Leu68 and Glu74) and the value of C_{ij} ranged from –1 (fully anti-correlated, the distance between the loop residue and active site gorge residue decreases while gorge diameter increases) to 1 (fully correlated, the distance between the loop residue and active site gorge residue increases while gorge diameter increases). The result of the matrix involving Glu74 is shown in Table 3-1. When combining the first rows of all three matrixes, which are the correlations between distance vectors and the gorge diameter, Table 3-2 is a direct measurement of the correlation between loop residue motions and gorge size.

CHAPTER 4 MAMMALIAN CARBOXYLESTERASES & ENZYME KINETICS OF PNBCE

4.1 ABSTRACT

Carboxylesterases (CEs) are ubiquitous enzymes that metabolize, among other compounds, the anticancer agent CPT-11. In this report, we used a bacterial enzyme, pnbCE as a model, to study the potential role of different structural components on enzyme metabolism. We identified two flexible loops, coil_5 and coil_21, in pnbCE using MD simulation and normal mode analysis. These loop regions are found to be highly variable among members of the α/β hydrolases fold superfamily. We generated loop-deleted mutants to assess the effects of such mutations on catalysis. Our results indicate that the removal of either one of the two surface loops (coil_5 and coil_21) dramatically reduces the k_{cat} of the reaction. This effect suggests the two loops potentially act like a 'lid' that protects the active site gorge by preventing substrates from diffusing out of the active site gorge. Additionally, we performed normal mode analysis on mammalian CEs and the results suggest their global motions are localized in the α-helices between αD and αE of the catalytic domain. The constraint of the hinge regions between these α-helices may explain the effect of previous mutation study on the side door residue Leu362. Overall, we propose that these highly flexible loops regions play an important role in catalysis by regulating the access of water or substrate to the active site.

4.2 INTRODUCTION

Mammalian Carboxylesterases (CEs; EC 3.1.1.1) are a family of ubiquitous proteins that can be found in the endoplasmic reticulum (ER) and cytosol of many tissues in mammals. CEs cleave carboxylic esters into the corresponding alcohol and carboxylic acid. These enzymes are capable of catalyzing the hydrolysis of ester- and amide-containing substrates with a variety of chemical structures, including the activation of prodrugs such as anti-cancer drug CPT-11 and capecitabine (*3*) (*104*). They may also be involved in metabolizing illicit street drugs such as heroin and cocaine (*105*). CEs can also hydrolyze compounds such as short- and long-chain acyl-glycerols, long-chain acyl-CoA esters, and long-chain acylcarnitine. Carboxylesterases are responsible for the detoxification of xenobiotics across various species.

Although the amino acid sequences of many CEs have been established using molecular cloning methods, according to the Ester database (http://bioweb.ensam.inra.fr/ESTHER/general?what=index), currently only 20 out 1394 CE sequences have crystal structures. Furthermore, these 20 structures are from 4 CE isozymes: human liver CE (hCE1), rabbit CE (rCE) and two bacterial CE (pnbCE and Est55). The remaining structures are either from different mutants or the same enzyme being crystallized with different inhibitors.

CEs are important in activating anti-cancer prodrug CPT-11. Both hCE1 and hiCE are known to convert CPT-11 into its active form, SN-38, a topoisomerase I inhibitor. However, hiCE in the small intestine is much more efficient at converting CPT-11 than hCE1 (*14*). Thus, the overproduction of SN-38 in the small intestine causes tissue damage to this organ and contributes to the delayed diarrhea in patients, a side-effect that often requires hospitalization (*14*, *21*).

We are specifically interested in development of selective inhibitors of a CE isozyme,

human intestine CE (hiCE), with the purpose of reducing or eliminating the CPT-11 activation in the small intestine (*25*). Since the crystal structure of hiCE is not available, we developed a homology model of hiCE and compared the dynamic features of four CE isozymes, hCE1, hiCE (homology model), rCE and pnbCE, as well as using site-specific mutation of pnbCE as a model system in an attempt to identify the structural features responsible for substrate specificity.

Despite being a bacterial model, as shown in Figure 4-1 and 4-2, pnbCE exhibits close sequence conservation and structural homology with mammalian CEs (*12*). The pnbCE has comparable activities with rabbit CE, which has the highest CPT-11 conversion rate among mammalian CEs toward the hydrolysis of o-NPA and CPT-11. Both enzymes showed reduced catalytic activities when corresponding amino acid residues were mutated (*12*). Since pnbCE is a bacterial model, it can be easily modified, expressed and purified in *E. coli* expression systems. Therefore, pnbCE makes an excellent model system for studying the structure-function relationship of human CEs. The information obtained from this study could ultimately be applied to the development of specific inhibitors of hiCE.

4.3 RESULTS

4.3.1 PROTEIN STRUCTURE MODELS

A total of 3 crystal structures, hCE1 (pdb: 1MX5) (*9*), rCE (1K4Y) (*13*) and pnbCE (1QE3) (*10*), were used in this study. The missing loops, residues 66-74 and 414-420 of pnbCE, and residues 355-370 and 450-466 of rCE, were inserted using Modeller (*106*). The structures of the deletion and insertion pnbCE mutants were modeled using Modeller as well. The final models of pnbCE mutants were evaluated with various tools including PROCHECK (*107*), VERIFY3D (*108*) and ERRAT2 (*109*) analysis. The overall quality factor of the wt- (1QE3), [coil_5]del- and

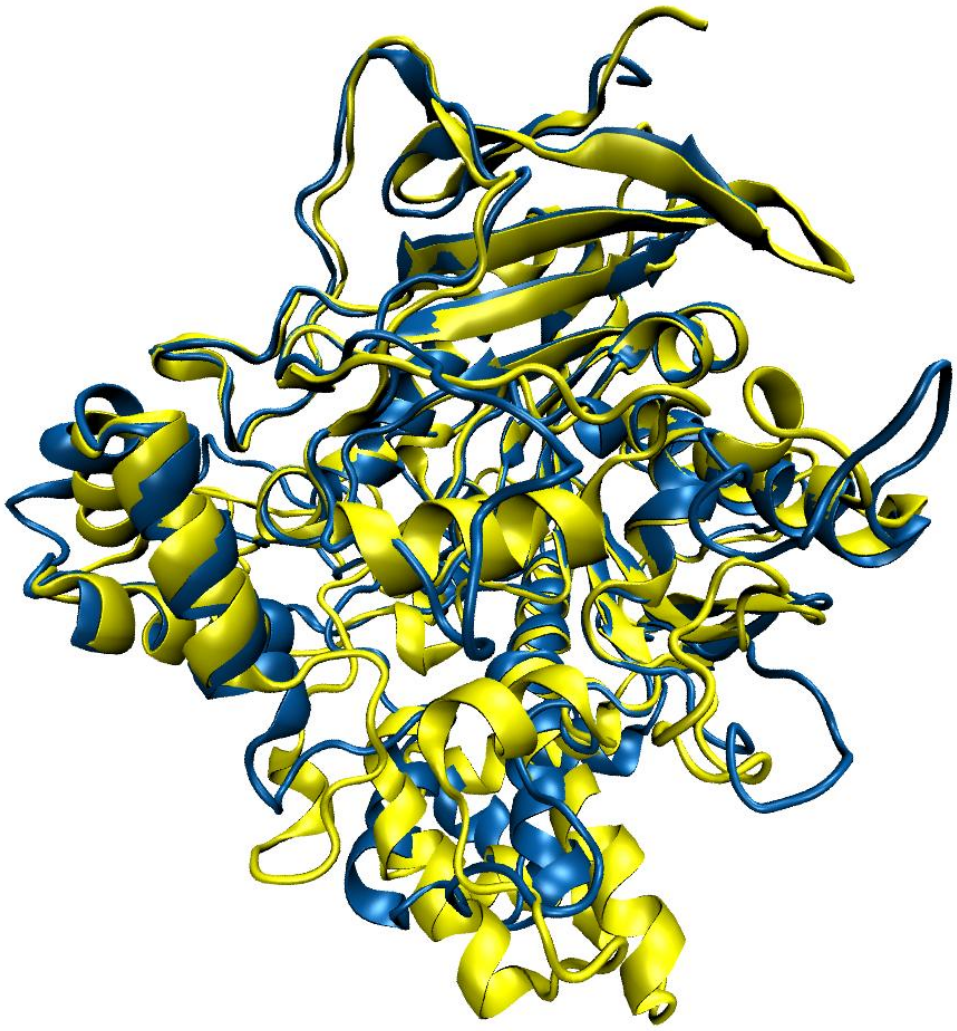


Figure 4-1. The alignment of x-ray crystal structures of pnbCE (PDB code: 1QE3) and hCE1 (1MX5)
Blue cartoon representation structure is pnbCE and yellow structure is hCE1. They both belong to the α/β fold hydrolases superfamily.

The alignment of the relevant CEs are given below.

```
rCE     MWLCALALASLAACTA---------WGHPSAPPVVDTVHGKVLGKFVSLEGFAQPVAVFL 51
hCE1    MWLRAFILATLSASAA---------WGHPSSPPVVDTVHGKVLGKFVSLEGFAQPVAIFL 51
hiCE    MRLHRLR-ARLSAVACGLLLLLVRGQGQDSASPIRTTHTGQVLGSLVHVKGANAGVQTFL 59
pnbCE   M----------------------------THQIVTTQYGKVKG------TTENGVHKWK 25
                                     :  :  *  *:* *          *  :

rCE     GVPFAKPPLGSLRFAPPQPAESWSHVKNTTSYPPMCSQDAVSGHMLSELFTNRKENIPLK 111
hCE1    GIPFAKPPLGPLRFTPPQPAEPWSFVKNATSYPPMCTQDPKAGQLLSELFTNRKENIPLK 111
hiCE    GIPFAKPPLGPLRFAPPEPPESWSGVRDGTTHPAMCLQDLTA--VESEFLSQFNMTFPSD 117
pnbCE   GIPYAKPPVGQWRFKAPEPPEVWEDVLDATAYGPICPQ-------PSDLLSLSYTELPRQ 78
        *:*:****:*  ** .*:*.* *. * : *:: .:* *        *::::     :* .

rCE     -FSEDCLYLNIYTPADLTKRGRLPVMVWIHGGGLMVGGAS--TYDGLALSAHENVVVVTI 168
hCE1    -LSEDCLYLNIYTPADLTKKNRLPVMVWIHGGGLMVGAAS--TYDGLALAAHENVVVVTI 168
hiCE    SMSEDCLYLSIYTPAHSHEGSNLPVMVWIHGGALVFGMAS--LYDGSMLAALENVVVVII 175
pnbCE   --SEDCLYVNVFAPDTPSQ--NLPVMVWIHGGAFYLGAGSEPLYDGSKLAAQGEVIVVTL 134
          ******:.:::*    :  .**********.: .* .*   ***  *:*  :*:** :

rCE     QYRLGIWGFFSTG--DEHSRGNWGHLDQVAALRWVQDNIANFGGDPGSVTIFGESAGGQS 226
hCE1    QYRLGIWGFFSTG--DEHSRGNWGHLDQVAALRWVQDNIASFGGNPGSVTIFGESAGGES 226
hiCE    QYRLGVLGFFSTG--DKHATGNWGYLDQVAALRWVQQNIAHFGGNPDRVTIFGESAGGTS 233
pnbCE   NYRLGPFGFLHLSSFDEAYSDNLGLLDQAAALKWVRENISAFGGDPDNVTVFGESAGGMS 194
        :****  **:  .  *:   .* * ***.***:**::**: ***:*. **:******* *

rCE     VSILLLSPLTKNLFHRAISESGVA--LLSSLFRKN-TKSLAEKIAIEAGCKTTTSAVMVH 283
hCE1    VSVLVLSPLAKNLFHRAISESGVA--LTSVLVKKGDVKPLAEQIAITAGCKTTTSAVMVH 284
hiCE    VSSLVVSPISQGLFHGAIMESGVA--LLPGLIASS-ADVISTVVANLSACDQVDSEALVG 290
pnbCE   IAALLAMPAAKGLFQKAIMESGASRTMTKEQAASTAAAFLQVLGINESQLDRLHTVAAED 254
        :: *:  * ::.**: ** ***.:  :      .  .  :       :  .   : .

rCE     CLRQKTEEELMEVTLKMKFMALDLVGDPKENTAFLTTVIDGVLLPKAPAEILAEKKYNML 343
hCE1    CLRQKTEEELLETTLKMKFLSLDLQGDPRESQPLLGTVIDGMLLLKTPEELQAERNFHTV 344
hiCE    CLRGKSKEEILAINKPFKMIPG---------------VVDGVFLPRHPQELLASADFQPV 335
pnbCE   LLKAADQLRIAEKENIFQLFFQP--------------ALDPKTLPEEPEKSIAEGAASGI 300
         *:   : .:      ::::                 .:*   * . * :  *.     :
```

```
rCE     PYMVGINQQEFGWIIPMQMLGYPLSEGKLDQKTATELLWKSYPIVNVSKELTPVATEKYL 403
hCE1    PYMVGINKQEFGWLIPMQLMSYPLSEGQLDQKTAMSLLWKSYPLVCIAKELIPEATEKYL 404
hiCE    PSIVGVNNNEFGWLIPKVMRIYDTQK-EMDREASQAALQKMLTLLMLPPTFGDLLREEYI 394
pnbCE   PLLIGTTRDEG-------YLFFTPDSDVHSQETLDAALEY-----LLGKPLAEKAADLYP 348
        * ::* ..:*           :  ..   .:::    *        :   :     : *

rCE     GGTDDPVKKKDLFLDMLADLLFGVPSVNVARHHRDAGAPTYMYEYRYRPSFSSDMRPKTV 463
hCE1    GGTDDTVKKKDLFLDLIADVMFGVPSVIVARNHRDAGAPTYMYEFQYRPSFSSDMKPKTV 464
hiCE    GDNGDPQTLQAQFQEMMADSMFVIPALQVAHFQCSR-APVYFYEFQHQPSWLKNIRPPHM 453
pnbCE   -------RSLESQIHMMTDLLFWRPAVAYASAQSHY-APVWMYRFDWHP-----EKPPYN 395
                     .:::* :*  *::  *  :    **.::*.:  :*      :*

rCE     IGDHGDEIFSVLGAPFLK-----EGATEEEIKLSKMVMKYWANFARNGNPNGEGLPQWPA 518
hCE1    IGDHGDELFSVFGAPFLK-----EGASEEEIRLSKMVMKFWANFARNGNPNGEGLPHWPE 519
hiCE    KADHGDELPFVFRSFFGGN---YIKFTEEEEQLSRKMMKYWANFARNGNPNGEGLPHWPL 510
pnbCE   KAFHALELPFVFGNLDGLERMAKAEITDEVKQLSHTIQSAWITFAKTGNPSTEAVN-WPA 454
         . *. *:  *:             ::*  :**: : . * .**:.***. *.:  **

rCE     YDYKEGYLQIGATTQAAQKLKDKEVAFWTELWAKEAAR--PRETEHIEL 565
hCE1    YNQKEGYLQIGANTQAAQKLKDKEVAFWTNLFAKKAVEK-PPQTEHIEL 567
hiCE    FDQEEQYLQLNLQPAVGRALKAHRLQFWKKALPQKIQELEEPEERHTEL 559
pnbCE   YHEET----------RETVILDSEITIENDPESEKRQKLFPSKGE---- 489
            :. :              :   .: : ..  .::  .    : .
```

Figure 4-2. The sequence alignment of rabbit CE, human liver CE, human intestine CE, and a bacterial CE

Coil_5 of pnbCE and its corresponding loops in mammalian CEs are highlighted in yellow and Coil_21 of pnbCE and its corresponding loops in mammalian CEs are highlighted in green.

[coil_21]del- pnbCEs are shown in Table 4-1. ERRAT is a program for verifying protein structures. Relatively high and consistent scores between wt- and mutant CEs suggest that the modeled mutant structures are reliable and accurate presentation of their protein structures. The 3D structures of these three mutants are almost identical with the wild-type enzyme, except for the mutated loop regions. Therefore, these models should be reliable representations of the mutant pnbCEs.

Table 4-1. The overall quality factors from the ERRAT2 server for wt- and mutant pnbCEs
The ERRAT2 score is a measurement of the accuracy of a protein structure. Our mutant CE structures have ERRAT2 scores that are fairly close to wt-pnbCE. It means these models are good representation of the mutant protein structures.

Protein	ERRAT2 score
wt-pnbCE	89.978
[coil_5]del-pnbCE	87.716
[coil_21]del-pnbCE	87.827

4.3.2 CONSERVATION STUDY OF PNBCE

The conservation study is done using the ConSurf server (*110*). The input sequence is the crystal structure of pnbCE (PDB: 1QE3). A sequence similarity search returns with 492 hits, including 350 unique protein sequences of known structure. These are the protein sequences similar to the pnbCE and they are probably also phylogenetically related to it. Finally, the conservation scores of every residue in pnbCE is calculated and colored with corresponding color. As shown in Figure 4-3, the warm color residues evolve slowly and are referred to as "conserved", while cold color residues evolve rapidly and are referred to as "variable". There are a few important results we can get from our conservation study. First of all, the α/β folding core of pnbCE, including the catalytic triad, is highly conserved. For example, among the catalytic triad,

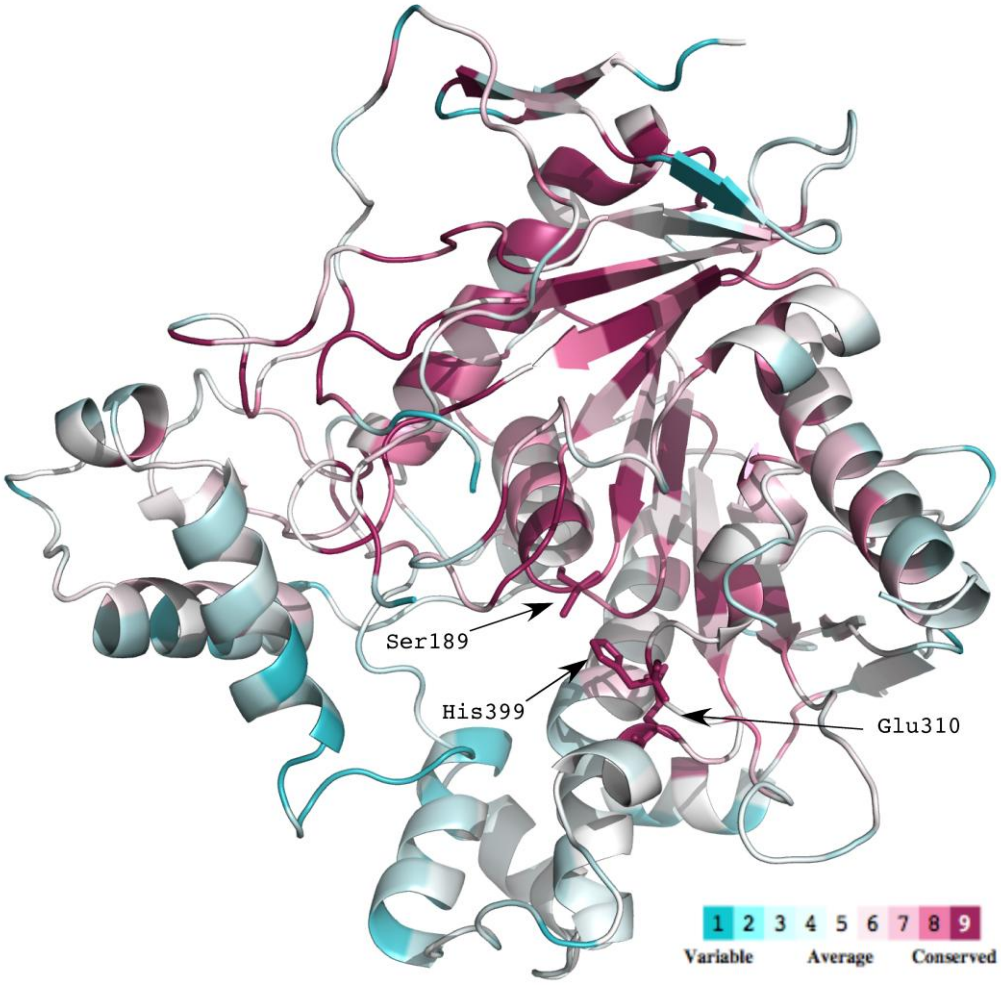


Figure 4-3. The conservation pattern obtained from the crystal structure of pnbCE
The 3-D structure of pnbCE is presented as a cartoon model, and colored according to the conservation scores. The color-coding bar shows the coloring scheme, with the warm color as conserved residues and cold color as variable residues. The catalytic domain of the enzyme, including the catalytic triad (Ser189, Glu310 and His399), is highly conserved while the regulatory domain of pnbCE is highly variable.

Ser189 and Glu310, each has only two variable residues among 350 sequences. His399 has no alternative residue and is the most conserved residue in the entire sequence. It is expected that the functional regions of the protein are highly conserved. That's the reason why these hydrolases are so diverse in sequence but be able to keep the same folding pattern and their common function – hydrolyze small organic molecules. Secondly, variability is just as important as conservation for reflecting protein functions. During the evolution of this particular enzyme, it adapts to chemical and biological envioronments in different organisms. In order for the same enzyme to perform slightly different tasks, such as selecting different substrates, it has to evolve residues around its functional regions to create this functional variability. One example is the surface loop regions, coil_5 and coil_21, as we talked about in the previous chapter, which is not just structurally flexible, but also highly variable in sequence. For instance, although coil_5 is missing in Figure 4-3 (because it's missing in the crystal structure), each residue in this loop has a conservation score of 1 or 2 (on a 1 to 9 scale with 1 being most variable). These loops may not be directly involved in the hydrolysis mechanism, but help the enzyme select different substrates based on their dynamic motions. This is also part of our goal in this study: to find out the structural features that are varible among the α/β hydrolases fold superfamily, as well as the structural features that are conserved.

Other highly variable regions in pnbCE structure are the α-helix bundles that consist of the lower side of the active site gorge. They are also the main components of the regulatory domain of pnbCE. Their dynamic motions correspond to the motions from residue 310 to residue 360, which we observed and will discuss in detail in our normal mode analysis. Similar structural features can be seen from other members of the α/β hydrolases fold superfamily. For instance, as one of the smallest member of the superfamily, Fusarium solani cutinase (FsCUT) has only 214 amino acids (comparing with pnbCE's 489) (*111*). Although its catalytic domain is conserved, it lacks the

majority part of the regulatory domain. Its catalytic triad is protected by two short hydrophobic loops, the loops 80 to 88 and 180 to 188. Both loops show greater anisotropy than other residues in FsCUT and the motions of these loops produces a breath-like movement that plays an important role in its catalysis (*111*). It is highly likely that the regulatory domains of CEs adopt a similar role as the hydrophobic loops of FsCUT. Our hypothesis is that these helices in the regulatory domains of pnbCE isolate the active site of the enzyme from external solvent and possible incorporated in the reaction mechanism by regulating the entrance of water or substrate into the active site.

4.3.3 NORMAL MODE ANALYSIS OF CES

Normal mode analysis (NMA) is the simplest way to attain large-scale, low-frequency motions of molecular structures (*81*) (*112*) (*113*). A total of 6 normal mode calculations were carried out, including the wild-type pnbCE (wt-), 2 loop deletion mutants ([coil_5]del- and [coil_21]del-), and 2 pnbCE mutants with known crystal structure (1C7J and 1C7I (*10*)). The latter two structures can be found in protein databank (PDB), and they both have multiple mutations comparing with the wt-pnbCE (1C7J has 11 mutations and 1C7I has 7 mutations). Some of their mutations stabilized the two loop regions that were missing in the wild-type (*10*), so both these two mutants have complete crystal structures.

The first five lowest-frequency vibrational normal modes for all wt- and mutant pnbCEs were generated by the ElNémo web server (*55*). Except for the [coil_21]del-pnbCE mutant, whose second lowest frequency mode (because the lowest frequency mode has a low collective value, which means this mode was localized in the termini) was used in analysis, all the other 5 calculations used their lowest frequency mode (mode 7). The root-mean-square-deviation (RMSD)

of their low frequency modes are plotted in Figure 4-4. The gray bars represent the RMSD of wt-pnbCE and the RMSD of the mutants are colored according to the color legend.

First of all, when we compare the wt-pnbCE with the two natural mutants (1C7J and 1C7I), it is not surprising that both loop regions (coil_5 and coil_21) have much lower RMSD values, just as their crystal structures suggested. On the other hand, it also verified the effectiveness of our wt-pnbCE structure. The wt-pnbCE has the two loops missing in its crystal structure and presented collective loop motions; while in contrast, the two mutants have stabilized loops in their crystal structures and were confirmed by the much constrained loop motions.

The low frequency modes of all the mutants have very similar fluctuation profile, which is expected since our structural models for these mutants have almost identical backbones except for the loop regions. Unlike the two natural mutations, [coil_21]del-pnbCE has a dominated loop motion around residues 66-76 similar to the wild-type, as its coil_5 is still intact. Similarly, the same motion as the wild-type was observed in the loop region 414-420 for the [coil_5]del-pnbCE. These two loop motions are structurally independent, which means deleting one loop doesn't affect the flexibility of the other loop.

The other most significant motions other than the loops are the α-helixes in residues 220-270 and residues 310-360, and they correspond to the α-helixes αD_1'-αD_3' (αD', the inserted α-helixes in front of the αD) and αE_1'-αE_3' (αE', the inserted α-helixes in front of the αE) of pnbCE in Figure 3-1, respectively. These α-helixes are not part of the α/β folding core, but are inserted into αD and αE during evolution of the enzyme, as evident of being highly variable in Figure 4-4 (the α-helixes located at the bottom of the structure). Interestingly, as part of the actual α/β folding core, αD (residues 288-293) and αE (residues 364-375) are located at the minima of the RMSD plot in Figure 4-4 and have extremely low flexibility compared with their neighboring

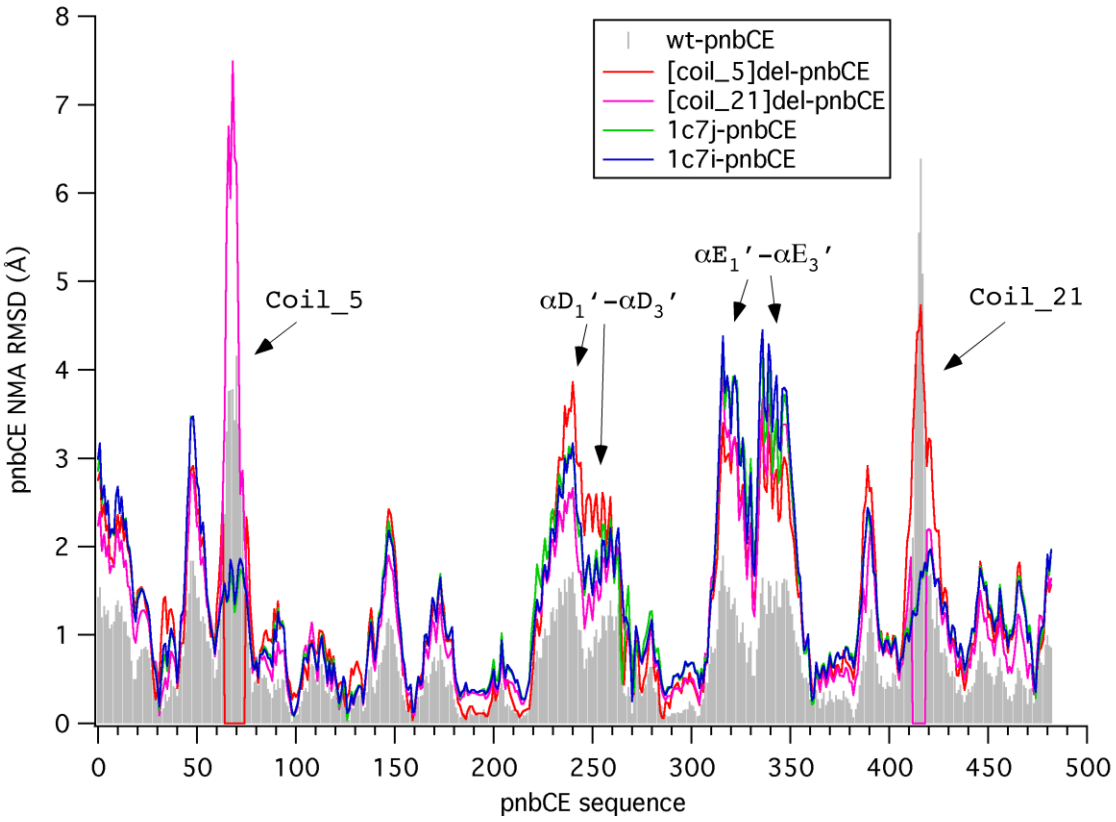


Figure 4-4. RMSD plots of the lowest-frequency normal mode of wt- and mutant pnbCEs
Here wt-pnbCE is plotted in gray bars and mutants are colored according to the legend.

regions. They act as “hinge regions” that holds the mobile parts of the enzyme together. These residues are structurally important and usually high conserved among the whole family of enzyme.

We also performed the same normal mode calculations on the mammalian CEs. Despite the fact their ability to convert large substrate, like CPT-11, differs dramatically, all 3 mammalian CEs exhibit almost identical low frequency normal modes. Although the loop motions are still there, they are not as dominant as the loops in the wt-pnbCE. Just like the natural pnbCE mutants, the dominated motions in these mammalian CEs are the α-helix inserts. The signification of these motions will be discussed later in the discussion section.

4.3.4 EXPRESSION AND PURIFICATION OF WILD-TYPE AND MUTANT PNBCES

The pnbA gene was inserted into the pTriEx-3 vector to create the pTriEx-3/pnbA plasmid DNA. Although we only expressed our protein in *E. coli* cells, this pTriEx-3 vector has the advantage of providing high levels of protein expression in both mammalian cell lines and in bacteria. All the mutants are generated by using the site-directed mutagnesis protocol and their sequences are verified by DNA sequencing. After the expression, pnbCEs were purified by using Isoelectric fousing (IEF) method based on their Isoelectric points (pI value). The concentration of the wt- and mutant pnbCEs are determined by Bradford protein assay. The SDS-PAGE gel of the IEF purified pnbCEs are shown in Figure 4-5. The purifed enzymes are used in the following circular dichroism spectroscopy study and enzymatic kinetic study. In Figure 4-6, it is shown the circular dichroism (CD) spectrums of the wild-type and mutant pnbCEs are similar, indicating the mutations didn’t change the overall folding of pnbCEs.

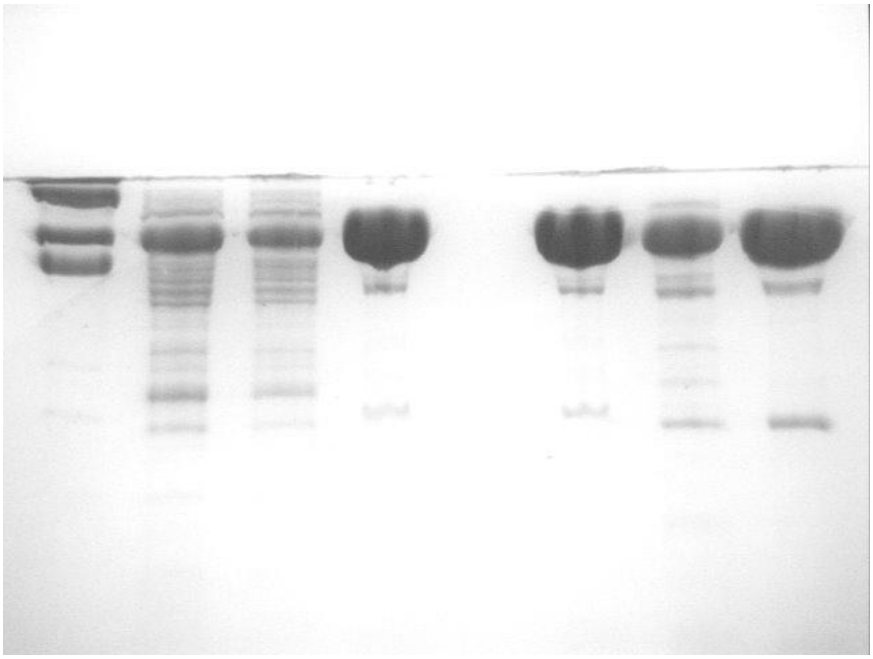


Figure 4-5. SDS-PAGE gel of the IEF purified pnbCEs
Approximately 10 μg of protein was loaded into each lane and following IEF purification, the gel was stained with Coomassie Blue. Lane 1 - protein standard; Lane 2 - After raw extraction; Lane 3 - After centrifuge and dialysis; Lane 4 - After IEF with ampholytes; Lane 5 - purified wt-pnbCE; Lane 6 - purified [coil_5]del-pnbCE; Lane 7 - purified [coil_21]del-pnbCE**.**

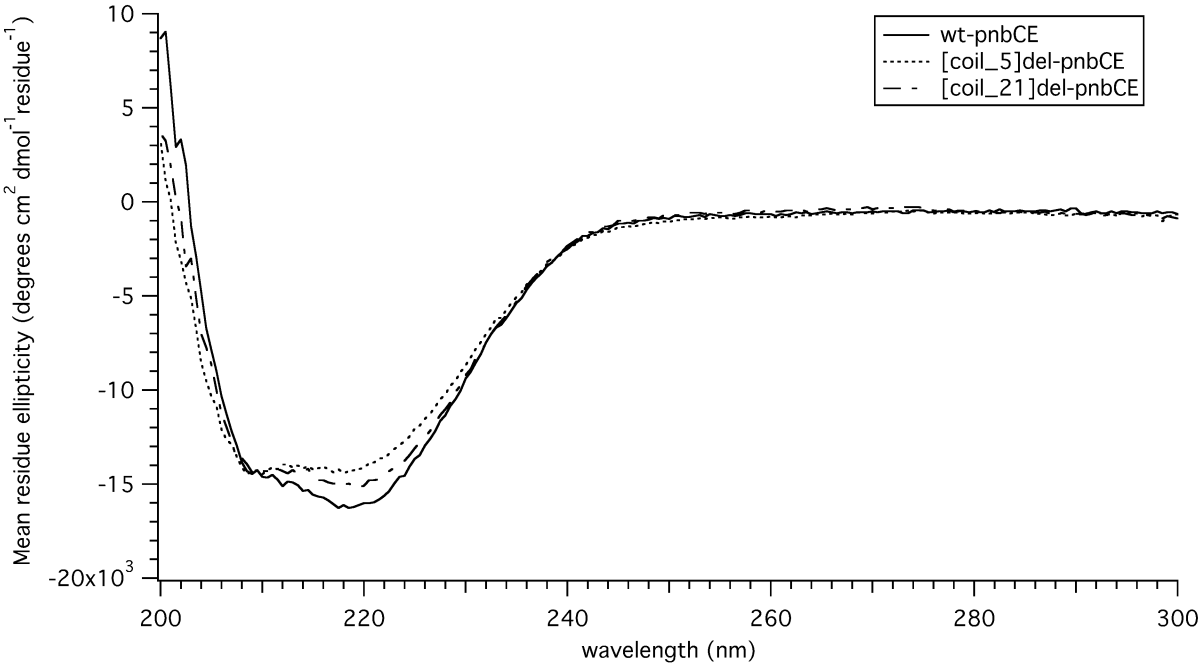


Figure 4-6. CD spectra of wt-, [coil_5]del- and [coil_21]del- pnbCEs

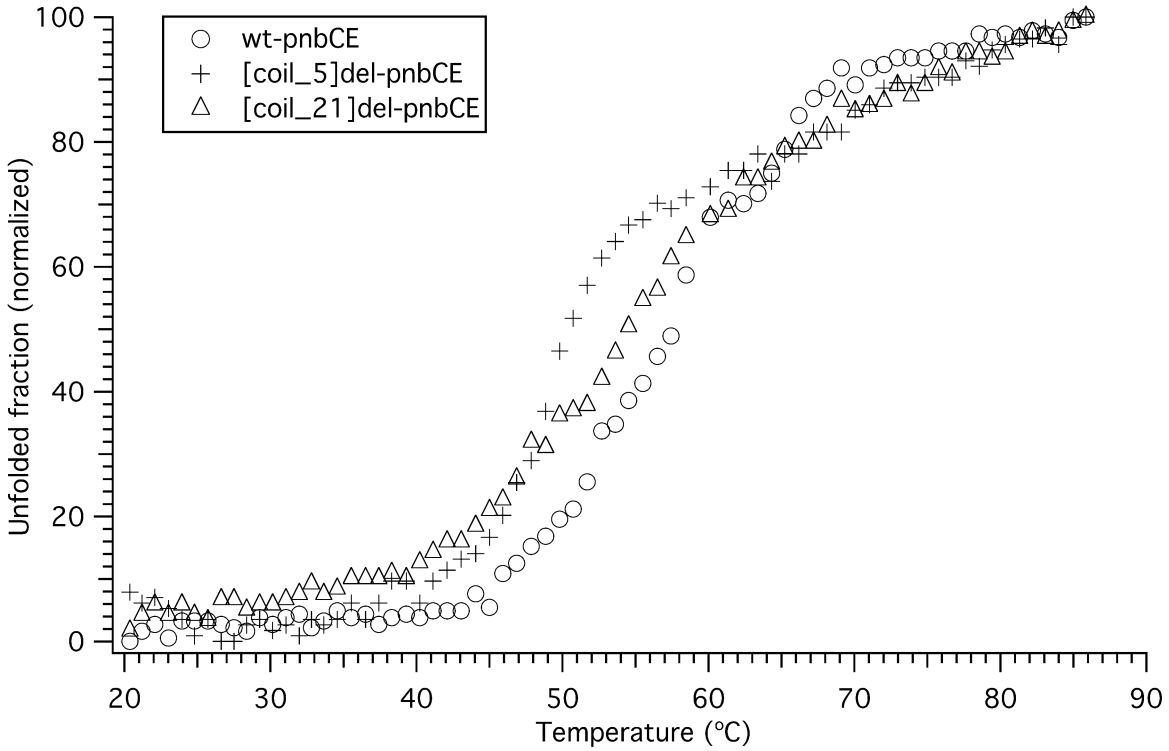


Figure 4-7. Temperature denaturation of wt-, [coil_5]del- and [coil_21]del-pnbCEs.

The melting temperature (T_m) of wt-pnbCE (Figure 4.7), 57.25 ℃, also matches to the literature value (*114*). The two mutants unfold at 51.04 ℃ for [coil_5]del-pnbCE and 54.45 ℃ for [coil_21]del-pnbCE.

We calibrated a Superose 12 10/300 column using protein standards and applied a sample of the purified CE. It indicated that pnbCE had a predicted molecular weight of ~54kDa and pnbCE purifies as a monomer.

4.3.5 ENZYME KINETICS OF WILD-TYPE AND MUTANT PNBCES

The ortho-nitrophenyl acetate (o-NPA) is a common substrate for measurement of the esterase activity, because its hydrolysis product o-nitrophenyl (o-NP) is a yellow solution that absorbs at 420 nm. Kinetic parameters for o-NPA were determined by assessing the conversion of o-NPA to o-NP. All 3 enzymes were tested, including the wt-, [coil_5]del- and [coil_21]del-pnbCEs. As shown in Figure 4-8, all the enzymes exhibited standard Michaelis-Menten kinetics. As shown in Table 4-2, modest increases in the Km value were observed for the [coil_5]del- and [coil_21]del- mutant pnbCE compared to that of the wild-type pnbCE. The *V*max and k_{cat} for the mutant CEs were 4.5-fold and 3-fold smaller than that for the wild-type enzyme, resulting in around 6-fold decrease in the catalytic efficiency of the loop-deleted mutant pnbCEs as compared to the wild-type enzyme.

4.4 DISCUSSION

Ultimately we are interested in the mechanism and efficiency of substrate hydrolysis, especially CPT-11, by mammalian CEs. The results presented in this study indicated that pnbCE can be used as a bacterial model for the analysis of specific mutations and their effects on drug

Table 4-2. Kinetic parameters of wt-, [coil_5]del- and [coil_21]del- pnbCE using o-NPA as a substrate

protein	K_m (μM)	V_{max} (μmol min^{-1} · mg^{-1} protein)	k_{cat} (s^{-1})	k_{cat}/K_m ($s^{-1}mM^{-1}$)
wt-pnbCE	1705 ± 159	149.9 ± 6.4	135.04	79.20
[coil_5]del-pnbCE	2338 ± 363	33.1 ± 2.6	29.82	12.75
[coil_21]del-pnbCE	3980 ± 1784	52.1 ± 14.1	46.94	11.79

K_m and V_{max} values are presented as mean ± S.E. Previous studies used the extinction coefficient value of 13,600 M^{-1} cm^{-1} (*2*) (*12*) (*115*) for o-NP at 420 nm. When using this value, our K_m and k_{cat} parameters of wt-pnbCE agree with reference value. However, it is reported that 5,220 M^{-1} cm^{-1} is the more accurate extinction coefficient for o-NP (*116*) (*117*) and the kinetic parameters in Table 4-2 are calculated using the new value.

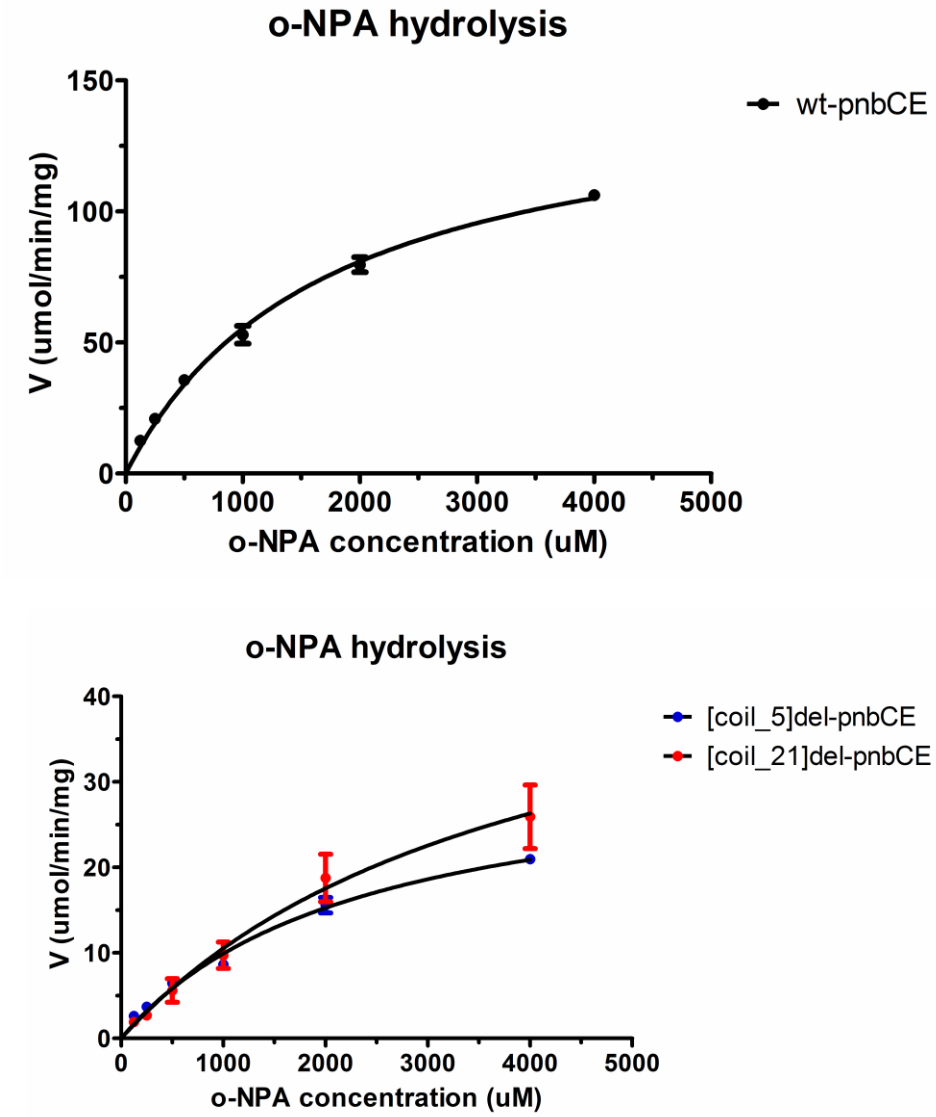


Figure 4-8. Non-linear regression curve fitting of o-NPA hydrolysis by wt-, [coil_5]del- and [coil_21]del- pnbCEs

catalysis. The roles of the surface dynamic loops gating the active site of different enzymes have been studied via various computational analyses (*118*) (*119*) (*120*). However, experimental evidence to support these proposed mechanisms are still very limited, especially for carboxylesterases.

Based on our computational models, our hypothesis is that these loops act as a gate that protects the active site and substrate from external solvent. We created specific mutations on those two flexible loops of pnbCE in order to eliminate/constrain their dynamic motions. Our current study demonstrates that removing these loop motions from the enzyme dramatically influences the kinetic parameters of catalysis, including rates of substrate binding (Km) and chemical bond associating/breaking (k_{cat}). As shown in Table 4-2, removing the loops reduced the rate of chemistry by around 4- fold. It clearly demonstrated the loops play an important role in the mechanism of hydrolysis. However, these two loops are located on the surface of the enzyme and they are away from the active site. Furthermore, our conservation study suggested that these loops are highly variable among members of the α/β hydrolases fold superfamily. Therefore, these loops were not expected to be directly involved in the catalysis (k_{cat}). It was recently reported that substituting certain amino acids in the hCE1 loops near the entrance of its active site with the corresponding residues from rCE resulted in an hCE1 variant that has dramatically different kinetic parameters from the wild-type enzyme (*121*). So it is not uncommon for some surface loops to be involved in the catalysis step in this family of enzyme. However, the detailed mechanism that these loops incorporated in the reaction is still unclear.

Since we demonstrated that the two loops (coil_5 and coil_21) have an open/close motion, one possible mechanism is that these two loops work together and act like a ‘lid’ that cover the active site gorge. Once the substrate binds near the active site gorge, the lid closes up and prevents

the substrate from diffusing out of the active site. By removing part of the lid, this highly regulated process becomes "leaky" and substrate tends to diffuse in both directions, in and out of the active site. Therefore lower the overall V_{max} of the catalysis. This 'lid' mechanism would explain the significant decreases of k_{cat} for the loop-deleted mutants comparing with wt-pnbCE.

Furthermore, the existence of a "side door" or "back door" were suggested on the crystal structures of acetylcholinesterases (*122*) (*17*) and rabbit carboxylesterases (*13*). Previous studies have also shown that mutations of the side door residue Leu362 in pnbCE significantly reduce the hydrolysis rates of both o-NPA and CPT-11 (*123*) (*16*). We performed normal mode analysis on mammalian CEs and the results are shown in Figure 4-9. The flexible α-helices in hCE1 are highlighted in Figure 4-9a and Figure 4-9b is the back view. The two blue residues are Val388 and Val424, which are the so-called "side door" residues in hCE1. They are located at the minimum RMSD value regions and likely act as hinge residues around the large α-helical insertions between αD and αE. These hinges are previously reviewed and thought to be important in facilitating the large-scale, slower lid motions that produce a catalytically competent state (*124*). It was also previously suggested that the Phe363, which is adjacent to the side door residue Leu362 in pnbCE, act as a "hinge' residue between the α-helices, which separate the regulatory and catalytic domains of CEs (*15*). As we demonstrated in our normal mode analysis (Figure 4-9c, d and e), these α-helices are the most flexible components of mammalian CEs and these global motions are most likely also biologically important. By making mutations on Leu362 could potentially constrain the rotational ability of the hinge region and therefore constrains the collective motions of the α-helices. Since previous experimental data suggested that the mutations at the side door residue Leu362 didn't affect the ionization of critical residues in the active site nor change the spatial orientation of the catalytic triad in the active site (*125*), our interpretation of the mutagenesis data

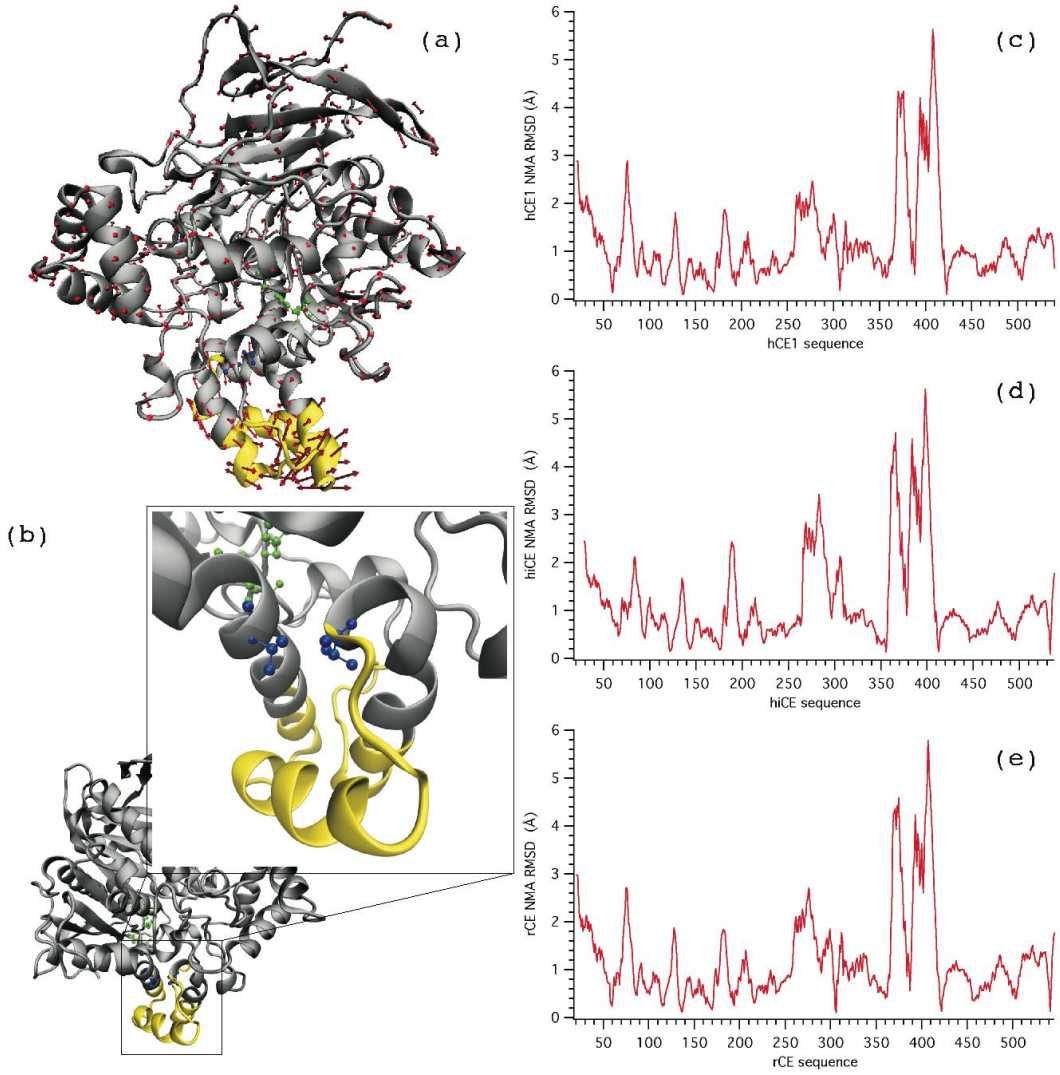


Figure 4-9. Normal mode analysis of mammalian CEs
(a) The front view of lowest-frequency normal mode of human liver CE (hCE1). The red arrow indicates the motion of the residues in this mode. Yellow highlighted region is the helixes that are most flexible in this mode. (b) The back view of the same structure as (a). The two blue residues are Val388 and Val424. They are the so-called "side door" residues in hCE1, which are located at the hinge of two flexible helixes. (c, d, e) The RMSD plots of the lowest-frequency modes for hCE1, hiCE and rCE, respectively.

on the side door residues Leu362 offers an alternative reason that could contributes to the effect of side door mutations on hydrolysis.

Overall we tested our hypothesis about the potential role of the two large surface loops in hydrolysis by making loop-deleted mutants and measuring their kinetic parameters. The mutants have shown significantly different kinetic parameters. Our results suggested that the loops are incorporated in the catalysis, possibly by preventing the substrate from diffusing out of the active site gorge. Our study also indicates that pnb CE is a valuable model for studying hydrolysis by mammalian CEs. Further experiments can be designed to test if this proposed mechanism about the loops is sequence-dependent. This can be done by swapping the loop sequences of pnbCE with the mammalian CEs. The result of this kinetic study should give us more insights about the role of these surface loops.

4.5 MATERIALS & METHODS

4.5.1 PROTEIN SEQUENCE CONSERVATION STUDY

The conservation study is done by using the ConSurf server (*110*). It is a useful tool that enables the identification of biologically important regions of a protein, of known crystal structure, based on its close sequence homologues.

The crystal structure of pnbCE (PDB: 1QE3) was used as the input file. The ConSurf server extracts the protein sequence information from the SEQRES record of the crystal structure or the ATOM record of other PDB file. Then it carries out a search for close homologous sequences using PSI-BLAST (*126*) in the UniProt database. The search uses the maximum and minimum sequence identity cutoffs set up by the user. It then performs the multiple sequence alignment (MSA) using the default MUSCLE (*127*) algorithm. The conservation score is

calculated using an empirical Bayesian method (*128*) based on the multiple sequence alignment result. Optionally, a phylogenetic tree can be built with the MSA result as well. Finally, the protein, with the conservation scores color-coded onto its surface, can be visualized using the Pymol (*129*) molecular visualization software. Some of the parameters used in the conservation study of pnbCE are listed in Table 4-3. To be noticed is that the max and min sequence identity cutoff used in this study is from 10% to 95% and it returns with 492 hits, including 350 unique protein sequences of known structure. Considering the total number of known crystal structures in the α/β hydrolases fold superfamily is a little over 900 sequences; this conservation study has a reliable sample pool that contains most of the phylogenetically related protein sequences.

Table 4-3. Parameters used in the conservation analysis of pnbCE

ConSurf parameter name	Parameter value
Protein	pnbCE
PDB code	1QE3
Multiple Sequence Alignment algorithm	MAFFT
Protein Database	UniProt
Method to calculate conservation scores	Bayesian
Model of substitution for proteins	JTT
Max %ID between sequences	95
Min %ID between sequences	10
PSI-BLAST hits	492
Unique sequences	350

4.5.2 NORMAL MODE ANALYSIS

Normal mode analysis (NMA) was performed on the ElNémo web server (*55*), which is a web interface to the Elastic Network Model that implements the 'rotation-translation-block' (RTB) approximation. The starting structures were described previously in section 4.3.1. The number of residues that are treated as a rigid body value (NRBL) was determined by the server as a function of protein size to optimize computation speed. A cutoff of 8 Å was used to identify elastic interactions. We are interested the 5 lowest non-zero frequency modes (modes 7-11) that were generated. The output of NMA is a series of trajectory files in Protein Data Bank (PDB) format, which include 11 structures by default: the original input structure and ten frames representing the motions between the two conformational extrema of the normal mode.

Some of the normal modes have a low degree-of-collectivity value less than 0.2, which meant only a few atoms within the N- or C-termini of the structure were involved in these motions. For the purpose of analyzing structural and functional relationship, these modes were ignored due to the lack of global conformational significance. The lowest-frequency non-localized mode for each enzyme was used for further analysis. A total of 9 normal mode analyses were performed in this study.

4.5.3 BACTERIAL STRAINS AND VECTORS

Host bacterial strain for plasmid preparation, XL-10 gold supercompetent cells, were purchased from Stratagene. The E. coli expression host strain, *Origami B(DE3)pLacI* and the expression vector pTriEx-3, was obtained from Novagen.

4.5.4 PNBA GENE ISOLATION

The pnbA gene encoding pnbCE was isolated from B. subtilis genomic DNA using PCR as described earlier (*12*). The pnbA coding sequence was inserted into the BamHI restriction site of pTriEx-3 vector, to create the pTriEx-3/pnbA plasmid DNA.

4.5.5 MUTAGENESIS

In general, all the mutations were generated by using the Phusion site-directed mutagenesis kit from Finnzymes. Using pTriEx-3/pnbA plasmid as a template, the [coil_5]-del pnbCE mutant was made by PCR using primer pair: 5'-(phosphate) CGCCAGTCCGAGGATTGCTTG -3' and 5'-(phosphate) AGACGGCTGCGGGCAAATAGAG. The [coil_21]-del pnbCE mutant was made using primer pair: 5'-(phosphate) ATTACGGATG AGGTGAAACA GCTTTCTC -3' and 5'-(phosphate) CAATCCGTCC AGATTTCCAA AGAC -3'. Linear PCR products were purified using QIAquick PCR purification kit (QIAGEN), and were ligated back into circular plasmid DNA. After transformed to XL-10 Gold supercompetent cells, both deletion mutants plasmid were purified using QIAfilter plasmid purification kit (QIAGEN). All three deletion mutants were confirmed by DNA sequencing (protein and DNA sequences are shown in Appendix A4 and A5, respectively).

4.5.6 PROTEIN EXPRESSION AND PURIFICATION

All 3 deletion mutants, along with wt-pnb plasmids were transformed into E. coli expression strain, Origami B cells. After 24 hours, a 5 mL overnight SOC medium containing the appropriate concentrations of antibiotics was grown from a single colony for each mutant and wild type. On the second day, the 5 mL overnight bacterial culture was transferred to 2x 500 mL SOC

medium with antibiotics. After 3~4 hours of cell growth, protein expression was induced with 1mM IPTG. Another 3~4 hours following induction, cells were harvested with centrifugation. Protein was extraced using 30 mL of Bug-Buster reagent (Novagen). Protein extracts were dialyzed against 10 mM HEPES buffer (pH 7.4) overnight before preceding to isoelectric focusing. wt-pnb and [coil_21]-del pnbCE mutant CEs were mixed with Bio-Lyte 3/5 ampholytes (Bio-Rad), while [coil_5]-del pnbCE mutant was mixed with Bio-Lyte 5/7 ampholytes. After 4~5 hours of focusing, 20 protein fractions were collected. CE activity was tested and the fractions with high levels of CE activity were pooled. Salt concentration was adjusted to 1 M using 4 M NaCl. The ampholytes were removed by overnight dialysis in 10 mM Hepes (pH 7.4) with 50 mM NaCl. Purity of the CEs were determined by SDS-PAGE gels (Figure 4-5). The molecular weight of each protein was determined by applying a protein sample to a calibrated Superose 12 10/300 chromatography column (GE Healthcare). Protein concentrations were determined by standard Bradford protein assay.

4.5.7 CE ASSAY

A spectrophotometric assay was used in determining CE activities. Generally, a constant amount of enzyme was incubated with 3 mM of o-NPA, as the substrate, in 50 mM HEPES buffer (pH 7.4). The changes in absorption per minute were measured spctrophotometrically at 420 nM using a Cary-4 UV-VIS spectrophotometer. The CE activities were expressed as micromoles of o-NP produced per minute per milligram of CE.

4.5.8 CIRCULAR DICHROISM SPECTROSCOPY

CD spectra for all three enzymes were acquired on a Model 202SF CD Spectrometer (Aviv) from 200 to 300 nm, with a 5 s averaging time in 5 mM sodium phosphate, 50 mM NaCl, pH 7.4. The measured delta absorbance is converted into the mean residue ellipticity unit (degrees cm^2 $dmol^{-1}$ $residue^{-1}$).

For the temperature-induced unfolding experiment, a 1-cm cuvette with a screw top was used to prevent evaporation of the solution at high temperatures. The temperature probe was inserted through the cuvette screw top to monitor temperature changes. Protein concentrations of 1 μM in 5 mM sodium phosphate, 50 mM NaCl, pH 7.4, were used for denaturation experiment. The unfolding measurement was performed within a temperature range of 15–85 ℃. The temperature ramp was set at 1 ℃/min, with data taken every degree. The equilibration time at each temperature was 30 s. The acquisition time was 5 s. The measurements were taken at 222 nm. The melting temperature is calculated based on the curve fitting of the data points into the empirical unfolding equation,

$$A = A_0 + \frac{A_f - A_0}{(1 + \frac{T}{T_m})^{\alpha}} \qquad \text{(Equation 4-1)}$$

Where A is the Δ Absorbance from CD spectra, T is the temperature of the probe, A_f and A_0 are the final and initial temperature, α is the curve coefficient and Tm is the melting temperature of the protein sample.

4.5.9 o-NPA KINETICS

In general, kinetic parameters were determined by incubating different concentrations of o-NPA with constant amounts of CEs. o-NPA solutions were diluted into ranges from 16 mM to

0.25 mM in 50 mM HEPES buffer (pH 7.4) with a 16 mM o-NPA stock. The conversion of o-NPA to o-NP was monitored at 420 nm for 30 s. The reaction velocities were calculated from the data points and the kinetic parameters and r^2 values are determined by using non-linear regression with Prism software (Graph Pad).

REFERENCES

REFERENCES

1. Cashman, J. R., Perotti, B. Y., Berkman, C. E., and Lin, J. (1996) Pharmacokinetics and molecular detoxication, *Environ Health Perspect 104 Suppl 1*, 23-40.
2. Hatfield, M. J., Tsurkan, L., Hyatt, J. L., Yu, X., Edwards, C. C., Hicks, L. D., Wadkins, R. M., and Potter, P. M. (2010) Biochemical and molecular analysis of carboxylesterase-mediated hydrolysis of cocaine and heroin, *Br J Pharmacol 160*, 1916-1928.
3. Danks, M. K., Morton, C. L., Krull, E. J., Cheshire, P. J., Richmond, L. B., Naeve, C. W., Pawlik, C. A., Houghton, P. J., and Potter, P. M. (1999) Comparison of activation of CPT-11 by rabbit and human carboxylesterases for use in enzyme/prodrug therapy, *Clin Cancer Res 5*, 917-924.
4. Alexson, S. E., Diczfalusy, M., Halldin, M., and Swedmark, S. (2002) Involvement of liver carboxylesterases in the in vitro metabolism of lidocaine, *Drug Metab Dispos 30*, 643-647.
5. Mentlein, R., Heiland, S., and Heymann, E. (1980) Simultaneous purification and comparative characterization of six serine hydrolases from rat liver microsomes, *Arch Biochem Biophys 200*, 547-559.
6. Mentlein, R., and Heymann, E. (1984) Hydrolysis of ester- and amide-type drugs by the purified isoenzymes of nonspecific carboxylesterase from rat liver, *Biochem Pharmacol 33*, 1243-1248.
7. Furihata, T., Hosokawa, M., Fujii, A., Derbel, M., Satoh, T., and Chiba, K. (2005) Dexamethasone-induced methylprednisolone hemisuccinate hydrolase: its identification as a member of the rat carboxylesterase 2 family and its unique existence in plasma, *Biochem Pharmacol 69*, 1287-1297.
8. Bencharit, S., Morton, C. L., Hyatt, J. L., Kuhn, P., Danks, M. K., Potter, P. M., and Redinbo, M. R. (2003) Crystal structure of the human carboxylesterase 1 complexed with the Alzheimer's drug tachrine: from binding promiscuity to selective inhibition, *Chem. & Biol. 10*, 341-349.
9. Bencharit, S., Morton, C. L., Xue, Y., Potter, P. M., and Redinbo, M. R. (2003) Structural basis of heroin and cocaine metabolism by a promiscuous human drug-processing enzyme, *Nat Struct Biol 10*, 349-356.
10. Spiller, B., Gershenson, A., Arnold, F. H., and Stevens, R. C. (1999) A structural view of evolutionary divergence, *Proc Natl Acad Sci U S A 96*, 12305-12310.
11. Potter, P. M., and Wadkins, R. M. (2006) Carboxylesterases - Ddetoxifying enzymes and targets for drug therapy, *Current Medicinal Chemistry 13*, 1045-1054.
12. Wierdl, M., Morton, C. L., Nguyen, N. K., Redinbo, M. R., and Potter, P. M. (2004) Molecular modeling of CPT-11 metabolism by carboxylesterases (CEs): use of pnb CE as a model, *Biochemistry 43*, 1874-1882.
13. Bencharit, S., Morton, C. L., Howard-Williams, E. L., Danks, M. K., Potter, P. M., and Redinbo, M. M. (2002) Structural insights into CPT-11 activation by mammalian carboxylesterases, *Nat. Struct. Biol. 9*, 337-342.
14. Wadkins, R. M., Morton, C. L., Weeks, J. K., Oliver, L., Wierdl, M., Danks, M. K., and Potter, P. M. (2001) Structural constraints affect the metabolism of 7-ethyl-10-[4-(1-piperidino)-1-piperidino]carbonyloxycamptothecin (CPT-11) by carboxylesterases., *Mol. Pharmacol. 60*, 355-362.
15. Bencharit, S., Edwards, C. C., Morton, C. L., Howard-Williams, E. L., Kuhn, P., Potter, P.

M., and Redinbo, M. R. (2006) Multisite promiscuity in the processing of endogenous substrates by human carboxylesterase 1, *J Mol Biol 363*, 201-214.
16. Streit, T. M., Borazjani, A., Lentz, S. E., Wierdl, M., Potter, P. M., Gwaltney, S. R., and Ross, M. K. (2008) Evaluation of the 'side door' in carboxylesterase-mediated catalysis and inhibition, *Biol Chem 389*, 149-162.
17. Bartolucci, C., Perola, E., Cellai, L., Brufani, M., and Lamba, D. (1999) "Back door" opening implied by the crystal structure of a carbamoylated acetylcholinesterase, *Biochemistry 38*, 5714-5719.
18. Terzyan, S., Wang, C. S., Downs, D., Hunter, B., and Zhang, X. C. (2000) Crystal structure of the catalytic domain of human bile salt activated lipase, *Protein Sci 9*, 1783-1790.
19. Tanizawa, A., Fujimori, A., Fujimori, Y., and Pommier, Y. (1994) Comparison of Topoisomerase I Inhibition, DNA Damage, and Cytotoxicity of Camptothecin Derivatives Presently in Clinical Trials, *J. Natl. Cancer Inst. 86*, 836-842.
20. Humerickhouse, R., Lohrbach, K., Li, L., Bosron, W. F., and Dolan, M. E. (2000) Characterization of CPT-11 hydrolysis by human liver carboxylesterase isoforms hCE-1 and hCE-2, *Cancer Res 60*, 1189-1192.
21. Khanna, R., Morton, C. L., Danks, M. K., and Potter, P. M. (2000) Proficient metabolism of irinotecan by a human intestinal carboxylesterase, *Cancer Res 60*, 4725-4728.
22. Cygler, M., Schrag, J. D., Sussman, J. L., Harel, M., Silman, I., Gentry, M. K., and Doctor, B. P. (1993) Relationship between Sequence Conservation and 3-Dimensional Structure in a Large Family of Esterases, Lipases, and Related Proteins, *Protein Science 2*, 366-382.
23. Satoh, T., and Hosokawa, M. (2006) Structure, function and regulation of carboxylesterases, *Chem Biol Interact 162*, 195-211.
24. Frey, P. A., Whitt, S. A., and Tobin, J. B. (1994) A low-barrier hydrogen bond in the catalytic triad of serine proteases, *Science 264*, 1927-1930.
25. Stoddard, S. V., YU, X., Potter, P. M., and Wadkins, R. M. (2010) In silico design and evaluation of carboxylesterase inhibitors, *Journal of Pesticide Science 35*, 240-249.
26. Lushington, G. H., Guo, J.-X., and Hurley, M. M. (2006) Acetylcholinesterase: molecular modeling with the whole toolkit, *Curr. Top. Med. Chem. 6*, 57-73.
27. Kua, J., Zhang, Y., and McCammon, J. A. (2002) Studying enzyme binding specificity in acetylcholinesterase using a combined molecular dynamics and multiple docking approach, *J. Amer. Chem. Soc. 124*, 8260-8267.
28. Heikinheimo, P., Goldman, A., Jeffries, C., and Ollis, D. L. (1999) Of barn owls and bankers: a lush variety of alpha/beta hydrolases, *Structure 7*, R141-146.
29. Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., Harel, M., Remington, S. J., Silman, I., Schrag, J., and et al. (1992) The alpha/beta hydrolase fold, *Protein Eng 5*, 197-211.
30. Hatfield, J. M., Wierdl, M., Wadkins, R. M., and Potter, P. M. (2008) Modifications of human carboxylesterase for improved prodrug activation, *Expert Opin Drug Metab Toxicol 4*, 1153-1165.
31. Adcock, S. A., and McCammon, J. A. (2006) Molecular dynamics: Survey of methods for simulating the activity of proteins, *Chemical Reviews 106*, 1589-1615.
32. Berendsen, H. J., and Hayward, S. (2000) Collective protein dynamics in relation to function, *Curr Opin Struct Biol 10*, 165-169.
33. Wang, W., Donini, O., Reyes, C. M., and Kollman, P. A. (2001) Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis, protein-ligand,

protein-protein, and protein-nucleic acid noncovalent interactions, *Annual Review of Biophysics and Biomolecular Structure 30*, 211-243.
34. Warshel, A. (2003) Computer simulations of enzyme catalysis: methods, progress, and insights, *Annu Rev Biophys Biomol Struct 32*, 425-443.
35. Case, D. A., Darden, T. A., T.E. Cheatham, I., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Merz, K. M., Wang, B., Pearlman, D. A., Crowley, M., Brozell, S., Tsui, V., Gohlke, H., Mongan, J., Hornak, V., Cui, G., Beroza, P., Schafmeister, C., Caldwell, J. W., Ross, W. S., and Kollman, P. A. (2004) AMBER 8, University of California, San Francisco.
36. Seno, Y., and Go, N. (1990) Deoxymyoglobin Studied by the Conformational Normal Mode Analysis .1. Dynamics of Globin and the Heme-Globin Interaction, *Journal of Molecular Biology 216*, 95-109.
37. Seno, Y., and Go, N. (1990) Deoxymyoglobin Studied by the Conformational Normal Mode Analysis .2. The Conformational Change Upon Oxygenation, *Journal of Molecular Biology 216*, 111-126.
38. Perahia, D., and Mouawad, L. (1995) Computation of Low-Frequency Normal-Modes in Macromolecules - Improvements to the Method of Diagonalization in a Mixed Basis and Application to Hemoglobin, *Comput Chem 19*, 241-246.
39. Hayward, S., Kitao, A., and Berendsen, H. J. C. (1997) Model-free methods of analyzing domain motions in proteins from simulation: A comparison of normal mode analysis and molecular dynamics simulation of lysozyme, *Proteins-Structure Function and Genetics 27*, 425-437.
40. Hinsen, K. (1998) Analysis of domain motions by approximate normal mode calculations, *Proteins-Structure Function and Genetics 33*, 417-429.
41. Hinsen, K., Thomas, A., and Field, M. J. (1999) Analysis of domain motions in large proteins, *Proteins-Structure Function and Genetics 34*, 369-382.
42. Bahar, I., Atilgan, A. R., and Erman, B. (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential, *Folding & Design 2*, 173-181.
43. Tama, F., Gadea, F. X., Marques, O., and Sanejouand, Y. H. (2000) Building-block approach for determining low-frequency normal modes of macromolecules, *Proteins-Structure Function and Genetics 41*, 1-7.
44. Li, G. H., and Cui, Q. (2002) A coarse-grained normal mode approach for macromolecules: An efficient implementation and application to Ca2+-ATPase, *Biophysical Journal 83*, 2457-2474.
45. Kurkcuoglu, O., Jernigan, R. L., and Doruker, P. (2004) Mixed levels of coarse-graining of large proteins using elastic network model succeeds in extracting the slowest motions, *Polymer 45*, 649-657.
46. Tirion, M. M. (1996) Large amplitude elastic motions in proteins from a single-parameter, atomic analysis, *Physical Review Letters 77*, 1905-1908.
47. Skjaerven, L., Hollup, S. M., and Reuter, N. (2009) Normal mode analysis for proteins, *J Mol Struc-Theochem 898*, 42-48.
48. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., K.M. Merz, J., Ferguson, D. M., Spellmeyer, D. C., T. Fox, J. W. C., and Kollman, P. A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules., *Journal of American Chemical Society 117*, 5179-5197.
49. Brooks, M. E. (1983) Validity of monitoring nocturnal penile tumescence for a single

night, *Urol Res 11*, 187-189.
50. Marques, O., and Sanejouand, Y. H. (1995) Hinge-bending motion in citrate synthase arising from normal mode calculations, *Proteins 23*, 557-560.
51. Tama, F., and Sanejouand, Y. H. (2001) Conformational change of proteins arising from normal mode calculations, *Protein Eng 14*, 1-6.
52. Bahar, I., Atilgan, A. R., and Erman, B. (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential, *Fold Des 2*, 173-181.
53. Haliloglu, T., Bahar, I., and Erman, B. (1997) Gaussian dynamics of folded proteins, *Physical Review Letters 79*, 3090-3093.
54. Bahar, I., Atilgan, A. R., Demirel, M. C., and Erman, B. (1998) Vibrational dynamics of folded proteins: Significance of slow and fast motions in relation to function and stability, *Physical Review Letters 80*, 2733-2736.
55. Suhre, K., and Sanejouand, Y. H. (2004) ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement, *Nucleic Acids Research 32*, W610-W614.
56. Hutchison, C. A., 3rd, Phillips, S., Edgell, M. H., Gillam, S., Jahnke, P., and Smith, M. (1978) Mutagenesis at a specific position in a DNA sequence, *J Biol Chem 253*, 6551-6560.
57. Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. (1986) Specific Enzymatic Amplification of DNA Invitro - the Polymerase Chain-Reaction, *Cold Spring Harb Sym 51*, 263-273.
58. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., and Arnheim, N. (1985) Enzymatic Amplification of Beta-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle-Cell Anemia, *Science 230*, 1350-1354.
59. SantaLucia, J., Jr. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics, *Proc Natl Acad Sci U S A 95*, 1460-1465.
60. Meselson, M., and Yuan, R. (1968) DNA restriction enzyme from E. coli, *Nature 217*, 1110-1114.
61. Wilson, G. G., and Murray, N. E. (1991) Restriction and modification systems, *Annu Rev Genet 25*, 585-627.
62. Masamune, Y., and Richardson, C. C. (1968) Enzymatic breakage and joining of deoxyribonucleic acid. IV. DNA synthesis in E. coli infected with ligase-negative mutants of phage T4, *Proc Natl Acad Sci U S A 61*, 1328-1335.
63. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal Biochem 72*, 248-254.
64. Compton, S. J., and Jones, C. G. (1985) Mechanism of dye response and interference in the Bradford protein assay, *Anal Biochem 151*, 369-374.
65. Spector, T. (1978) Refinement of the coomassie blue method of protein quantitation. A simple and linear spectrophotometric assay for less than or equal to 0.5 to 50 microgram of protein, *Anal Biochem 86*, 142-146.
66. Satoh, T., and Hosokawa, M. (1998) The mammalian carboxylesterases: from molecules to functions, *Annu Rev Pharmacol Toxicol 38*, 257-288.
67. Hicks, L. D., Hyatt, J. L., Moak, T., Edwards, C. C., Tsurkan, L., Wierdl, M., Ferreira, A. M., Wadkins, R. M., and Potter, P. M. (2007) Analysis of the inhibition of mammalian carboxylesterases by novel fluorobenzoins and fluorobenzils, *Bioorg Med Chem 15*,

3801-3817.
68. Hicks, L. D., Hyatt, J. L., Stoddard, S., Tsurkan, L., Edwards, C. C., Wadkins, R. M., and Potter, P. M. (2009) Improved, selective, human intestinal carboxylesterase inhibitors designed to modulate 7-ethyl-10-[4-(1-piperidino)-1-piperidino]carbonyloxycamptothecin (Irinotecan; CPT-11) toxicity, *J Med Chem 52*, 3742-3752.
69. Hyatt, J. L., Moak, T., Hatfield, M. J., Tsurkan, L., Edwards, C. C., Wierdl, M., Danks, M. K., Wadkins, R. M., and Potter, P. M. (2007) Selective inhibition of carboxylesterases by isatins, indole-2,3-diones, *J Med Chem 50*, 1876-1885.
70. Hyatt, J. L., Stacy, V., Wadkins, R. M., Yoon, K. J., Wierdl, M., Edwards, C. C., Zeller, M., Hunter, A. D., Danks, M. K., Crundwell, G., and Potter, P. M. (2005) Inhibition of carboxylesterases by benzil (diphenylethane-1,2-dione) and heterocyclic analogues is dependent upon the aromaticity of the ring and the flexibility of the dione moiety, *J Med Chem 48*, 5543-5550.
71. Wadkins, R. M., Hyatt, J. L., Edwards, C. C., Tsurkan, L., Redinbo, M. R., Wheelock, C. E., Jones, P. D., Hammock, B. D., and Potter, P. M. (2007) Analysis of mammalian carboxylesterase inhibition by trifluoromethylketone-containing compounds, *Mol Pharmacol 71*, 713-723.
72. Wadkins, R. M., Hyatt, J. L., Wei, X., Yoon, K. J., Wierdl, M., Edwards, C. C., Morton, C. L., Obenauer, J. C., Damodaran, K., Beroza, P., Danks, M. K., and Potter, P. M. (2005) Identification and characterization of novel benzil (diphenylethane-1,2-dione) analogues as inhibitors of mammalian carboxylesterases, *J Med Chem 48*, 2906-2915.
73. Wadkins, R. M., Hyatt, J. L., Yoon, K. J., Morton, C. L., Lee, R. E., Damodaran, K., Beroza, P., Danks, M. K., and Potter, P. M. (2004) Discovery of novel selective inhibitors of human intestinal carboxylesterase for the amelioration of irinotecan-induced diarrhea: synthesis, quantitative structure-activity relationship analysis, and biological activity, *Mol Pharmacol 65*, 1336-1343.
74. Nagel, Z. D., and Klinman, J. P. (2009) A 21st century revisionist's view at a turning point in enzymology, *Nat Chem Biol 5*, 543-550.
75. Shen, T., Tai, K., Henchman, R. H., and McCammon, J. A. (2002) Molecular dynamics of acetylcholinesterase, *Acc Chem Res 35*, 332-340.
76. Wintrode, P. L., Zhang, D., Vaidehi, N., Arnold, F. H., and Goddard, W. A., III. (2003) Protein Dynamics in a Family of Laboratory Evolved Thermophilic Enzymes, *J. Mol. Biol. 327*, 745-757.
77. Vistoli, G., Pedretti, A., Mazzolari, A., and Testa, B. (2010) In silico prediction of human carboxylesterase-1 (hCES1) metabolism combining docking analyses and MD simulations, *Bioorganic & Medicinal Chemistry 18*, 320-329.
78. Hammes-Schiffer, S., and Benkovic, S. J. (2006) Relating protein motion to catalysis, *Annu Rev Biochem 75*, 519-541.
79. Schwartz, S. D., and Schramm, V. L. (2009) Enzymatic transition states and dynamic motion in barrier crossing, *Nat Chem Biol 5*, 551-558.
80. Bencharit, S., Morton, C. L., Hyatt, J. L., Kuhn, P., Danks, M. K., Potter, P. M., and Redinbo, M. R. (2003) Crystal structure of human carboxylesterase 1 complexed with the Alzheimer's drug tacrine: from binding promiscuity to selective inhibition, *Chem Biol 10*, 341-349.
81. Tama, F. (2003) Normal mode analysis with simplified models to investigate the global dynamics of biological systems, *Protein and Peptide Letters 10*, 119-132.

82. Bahar, I., Atligan, A. R., and Erman, B. (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential, *Folding & Design 2*, 173-181.
83. Tirion, M. M. (1996) Large amplitude elastic motions in proteins from a single-parameter, atomic analysis., *Phys Rev Lett 77*, 1905-1908.
84. McCammon, J. A., and Harvey, S. C. (1987) *Dynamics of proteins and nucleic acids*, Cambridge University Press, Cambridge [Cambridgeshire] ; New York.
85. Cygler, M., Schrag, J. D., Sussman, J. L., Harel, M., Silman, I., Gentry, M. K., and Doctor, B. P. (1993) Relationship between sequence conservation and three-dimensional structure in a large family of esterases, lipases, and related proteins, *Protein Sci 2*, 366-382.
86. Tai, K., Shen, T. Y., Borjesson, U., Philippopoulos, M., and McCammon, J. A. (2001) Analysis of a 10-ns molecular dynamics simulation of mouse acetylcholinesterase, *Biophysical Journal 81*, 715-724.
87. Shi, J. X., Boyd, A. E., Radic, Z., and Taylor, P. (2001) Reversibly bound and covalently attached ligands induce conformational changes in the omega loop, Cys(69)-Cys(96), of mouse acetylcholinesterase, *Journal of Biological Chemistry 276*, 42196-42204.
88. Shi, J. X., Tai, K., McCammon, J. A., Taylor, P., and Johnson, D. A. (2003) Nanosecond dynamics of the mouse acetylcholinesterase Cys(69)-Cys(96) omega loop, *Journal of Biological Chemistry 278*, 30905-30911.
89. Hosea, N. A., Radic, Z., Tsigelny, I., Berman, H. A., Quinn, D. M., and Taylor, P. (1996) Aspartate 74 as a primary determinant in acetylcholinesterase governing specificity to cationic organophosphonates, *Biochemistry 35*, 10995-11004.
90. Szegletes, T., Mallender, W. D., Thomas, P. J., and Rosenberry, T. L. (1999) Substrate binding to the peripheral site of acetylcholinesterase initiates enzymatic catalysis. Substrate inhibition arises as a secondary effect, *Biochemistry 38*, 122-133.
91. Mallender, W. D., Szegletes, T., and Rosenberry, T. L. (2000) Acetylthiocholine binds to asp74 at the peripheral site of human acetylcholinesterase as the first step in the catalytic pathway, *Biochemistry 39*, 7753-7763.
92. Sali, A., and Blundell, T. L. (1993) Comparative Protein Modeling by Satisfaction of Spatial Restraints, *Journal of Molecular Biology 234*, 779-815.
93. Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham III, T. E., DeBolt, S., Ferguson, D., Seibel, G., and Kollman, P. (1995) AMBER, a Package of Computer Programs for Applying Molecular Mechanics, Normal Mode Analysis, Molecular Dynamics and Free Energy Calculations to Simulate the Structural and Energetic Properties of Molecules., *Comp. Phys. Commun. 91*, 1-41.
94. Frishman, D., and Argos, P. (1995) Knowledge-based protein secondary structure assignment, *Proteins 23*, 566-579.
95. Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., and Liang, J. (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues, *Nucleic Acids Res 34*, W116-118.
96. Edelsbrunner, H., Facello, M., and Liang, J. (1998) On the definition and the construction of pockets in macromolecules, *Discrete Applied Mathematics 88*, 83-102.
97. Lee, B., and Richards, F. M. (1971) The interpretation of protein structures: estimation of static accessibility, *J Mol Biol 55*, 379-400.
98. Connolly, M. L. (1983) Analytical Molecular-Surface Calculation, *Journal of Applied Crystallography 16*, 548-558.
99. Hawkins, G. D., Cramer, C. J., and Truhlar, D. G. (1996) Parametrized models of aqueous

free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium, *Journal of Physical Chemistry 100*, 19824-19839.
100. Morishita, T. (2000) Fluctuation formulas in molecular-dynamics simulations with the weak coupling heat bath, *Journal of Chemical Physics 113*, 2976-2982.
101. Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes., *Journal of Computional Physics 23*, 327-341.
102. Hinsen, K. (2000) The molecular modeling toolkit: A new approach to molecular simulations, *Journal of Computational Chemistry 21*, 79-85.
103. Tara, S., Helms, V., Straatsma, T. P., and McCammon, J. A. (1999) Molecular dynamics of mouse acetylcholinesterase complexed with huperzine A, *Biopolymers 50*, 347-359.
104. Tabata, T., Katoh, M., Tokudome, S., Nakajima, M., and Yokoi, T. (2004) Identification of the cytosolic carboxylesterase catalyzing the 5'-deoxy-5-fluorocytidine formation from capecitabine in human liver, *Drug Metab Dispos 32*, 1103-1110.
105. Hatfield, M. J., Tsurkan, L., Hyatt, J. L., Yu, X., Edwards, C. C., Hicks, L. D., Wadkins, R. M., and Potter, P. M. (2010) Biochemical and molecular analysis of carboxylesterase-mediated hydrolysis of cocaine and heroin, *Brit J Pharmacol 160*, 1916-1928.
106. Sali, A., and Blundell, T. L. (1993) Comparative Protein Modelling by Satisfaction of Spatial Restraints., *J. Mol. Biol. 234*, 779-815.
107. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR, *J Biomol NMR 8*, 477-486.
108. Bowie, J. U., Luthy, R., and Eisenberg, D. (1991) A method to identify protein sequences that fold into a known three-dimensional structure, *Science 253*, 164-170.
109. Colovos, C., and Yeates, T. O. (1993) Verification of protein structures: patterns of nonbonded atomic interactions, *Protein Sci 2*, 1511-1519.
110. Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids, *Nucleic Acids Res 38 Suppl*, W529-533.
111. Longhi, S., Czjzek, M., Lamzin, V., Nicolas, A., and Cambillau, C. (1997) Atomic resolution (1.0 A) crystal structure of Fusarium solani cutinase: stereochemical analysis, *J Mol Biol 268*, 779-799.
112. Bahar, I., and Rader, A. J. (2005) Coarse-grained normal mode analysis in structural biology, *Curr Opin Struct Biol 15*, 586-592.
113. Bahar, I., Lezon, T. R., Yang, L. W., and Eyal, E. (2010) Global dynamics of proteins: bridging between structure and function, *Annu Rev Biophys 39*, 23-42.
114. Reyes-Duarte, D., Polaina, J., Lopez-Cortes, N., Alcalde, M., Plou, F. J., Elborough, K., Ballesteros, A., Timmis, K. N., Golyshin, P. N., and Ferrer, M. (2005) Conversion of a carboxylesterase into a triacylglycerol lipase by a random mutation, *Angew Chem Int Ed Engl 44*, 7553-7557.
115. Danks, M. K., Morton, C. L., Pawlik, C. A., and Potter, P. M. (1998) Overexpression of a Rabbit Liver Carboxylesterase Sensitizes Human Tumor Cells to CPT-11., *Cancer Res. 58*, 20-22.
116. Soundararajan, M., Bailey, C. P., and Markwell, J. (2008) Use of a laboratory exercise on molar absorptivity to help students understand the authority of the primary literature,

Biochem Mol Biol Edu 36, 61-64.
117. L. L áng, J. S., G. Vars ányi, M. Vizesy. (1961) Absorption Spectra in the Ultraviolet and Visible Region., In *A Theoretical and Technical Introduction.*, p 413, Academic Press Inc., New York.
118. Amaro, R. E., Minh, D. D., Cheng, L. S., Lindstrom, W. M., Jr., Olson, A. J., Lin, J. H., Li, W. W., and McCammon, J. A. (2007) Remarkable loop flexibility in avian influenza N1 and its implications for antiviral drug design, *J Am Chem Soc 129*, 7764-7765.
119. Baron, R., and McCammon, J. A. (2007) Dynamics, hydration, and motional averaging of a loop-gated artificial protein cavity: the W191G mutant of cytochrome c peroxidase in water as revealed by molecular dynamics simulations, *Biochemistry 46*, 10629-10642.
120. Amaro, R. E., Cheng, X., Ivanov, I., Xu, D., and McCammon, J. A. (2009) Characterizing loop dynamics and ligand recognition in human- and avian-type influenza neuraminidases via generalized born molecular dynamics and end-point free energy calculations, *J Am Chem Soc 131*, 4702-4709.
121. Wierdl, M., Tsurkan, L., Hyatt, J. L., Edwards, C. C., Hatfield, M. J., Morton, C. L., Houghton, P. J., Danks, M. K., Redinbo, M. R., and Potter, P. M. (2008) An improved human carboxylesterase for enzyme/prodrug therapy with CPT-11, *Cancer Gene Ther 15*, 183-192.
122. Gilson, M. K., Straatsma, T. P., McCammon, J. A., Ripoli, D. R., Faerman, C. H., Axelsen, P. H., Silman, I., and Sussman, J. L. (1994) Open "Back Door" in a Molecular Dynamics Simulation of Acetylcholinesterase, *Science 263*, 1276-1278.
123. Wierdl, M., Morton, C. L., Nguyen, N. K., Redinbo, M. R., and Potter, P. M. (2004) Molecular modeling of CPT-11 metabolism by carboxylesterases (CEs): Use of pnb CE as a model., *Biochemistry 43*, 1874-1882.
124. Henzler-Wildman, K. A., Lei, M., Thai, V., Kerns, S. J., Karplus, M., and Kern, D. (2007) A hierarchy of timescales in protein dynamics is linked to enzyme catalysis, *Nature 450*, 913-916.
125. Streit, T. M., Borazjani, A., Lentz, S. E., Wierdl, M., Potter, P. M., Gwaltney, S. R., and Ross, M. K. (2008) Evaluation of the 'side door' in carboxylesterase-mediated catalysis and inhibition., *Biol Chem 389*, 149-162.
126. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res 25*, 3389-3402.
127. Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res 32*, 1792-1797.
128. Mayrose, I., Graur, D., Ben-Tal, N., and Pupko, T. (2004) Comparison of site-specific rate-inference methods for protein sequences: empirical Bayesian methods are superior, *Mol Biol Evol 21*, 1781-1791.
129. DeLano, W. L. (2008) The PyMOL Molecular Graphics System., DeLano Scientific LLC, Palo Alto, CA, USA.

APPENDIX

APPENDIX

A.1 PERL SCRIPTS

A.1.1 Gorge_file_process.pl

This script is used to generate the gorge fluctuation data in MD.

```
#!/usr/bin/perl -w
open outFile, ">pnbCE_Gorge_10ns.dat";
@files = <*.dat>;
foreach $file (@files) {
  $file =~ /^\D+(\d+).dat$/;
  $ns = $1 - 12;
  open inFile, $file;
  while (defined($line = <inFile>)) {
        chomp $line;
        $line =~ /^(\d+), (\S+)$/;
        $time = $ns + 0.0001*$1;
        $gorgeSize = 10*$2;
        print outFile "$time\t$gorgeSize\n";
        }
}
close inFile;
close outFile;
```

A.1.2 gorge_correlation.pl

This is script is used in analyzing the vector and gorge diameter correlation data.

```
#!/usr/bin/perl -w
open outFile_c, ">pnbCE_overall_correlation_2.dat";
print outFile_c "residue number\tcorrelation value\n";
@files = <*.rms>;
foreach $file (@files) {
        $file =~ /^\D+(\d+).rms$/;
        $CA = $1;
        open outFile_g, ">pnbCE_gorge_$CA.dat";
        print outFile_g "correlation calculation data between gorge diameter and RMSF of residue
$CA\n\n";
        print outFile_g "time(ns)\tRMS(A)\tGorge Diameter\tGorgeDiaRMSF(A)\n";
        open RMS, $file;
        $sumTop = $sumBottomRMS = $sumBotGorgeDiaRMSF = 0;
        open GorgeDia, "pnbCE_gorge_diameter.dat";
        while (defined($lineRMS = <RMS>)) {
```

```
                chomp $lineRMS;
                $lineRMS =~ /^\s*([0-9]*\.?[0-9]*)\s+([0-9]*\.?[0-9]*)$/;
                $RMS = $2;
                $lineDia = <GorgeDia>;
                chomp $lineDia;
                $lineDia =~ /^\s*([0-9]*\.?[0-9]*)\s+([0-9]*\.?[0-9]*)$/;
                $time = $1;
                $gorgeDia = 2*$2;
                $gorgeDiaRMSF = sqrt(($gorgeDia - 4.0522) * ($gorgeDia - 4.0522));
                $sumTop = ($RMS * $gorgeDiaRMSF) + $sumTop;
                $sumBotRMS = ($RMS * $RMS) + $sumBottomRMS;
                $sumBotGorgeDiaRMSF = ($gorgeDiaRMSF * $gorgeDiaRMSF) +
$sumBotGorgeDiaRMSF;
                print outFile_g "$time\t$RMS\t$gorgeDia\t$gorgeDiaRMSF\n";
        }
        $aveTop = $sumTop / 100000;
        $aveBotRMS = $sumBotRMS / 100000;
        $aveBotGorgeDiaRMSF = $sumBotGorgeDiaRMSF / 100000;
        $aveBottom = sqrt($aveBotRMS * $aveBotGorgeDiaRMSF);
        $correlation = $aveTop / $aveBottom;
        print outFile_c "$CA\t$correlation\n";
        close RMS;
        close outFile_g;
        close GorgeDia;
}
close outFile_c;
```

A.1.3 gorge_correlation.pl

This is script is used to generating RMSD data.

```
open outFile, ">>deviations_from_ave.ptraj";
$count = 1;
while ($count <= 489) {
        print outFile "rms reference out pnbCE_dev_$count.rms time 0.1 :$count\n";
        $count ++;
}
close outFile;
```

A.1.4 NMA_RMSD_process.pl

*This is a perl script that processes the .rmsmod files got from normal mode analysis (elnemo

server).*

```
$pi = 3.14159;
@files = <*.rmsmod>;
foreach $file (@files) {
	open inFile, $file;
	open outFile, ">$file.RMSD";
	while (defined($line = <inFile>)) {
		chomp $line;
		@fields = split (/\s+/, $line);
		if ($fields[4] eq "CA") {
			$RMSD = $fields[5] * 100;
			$BF = ($RMSD * $RMSD) * 8/3 * $pi * $pi;
			print outFile "$fields[3]\t$fields[2]\t$RMSD\t$BF\n";
		}
	}
	close inFile;
	close outFile;
}
```

A.1.5 pnbCE_arrow.pl

*This is a perl script that is used to process the pdb files from NMA. All the CA atom coordinates

are passed into a VMD tcl file.*

```
@files = <pnbCE_mode7.pdb>;
foreach $file (@files) {
	open inFile, $file;
	open model_1, ">$file.m1";
	open model_11, ">$file.m11";

	while (defined($line = <inFile>)) {
		chomp $line;
		@fields = split (/\s+/, $line);
		if ($fields[0] eq "MODEL" and $fields[1] == "1") {
				$line = <inFile>;
				$line = <inFile>;
				chomp $line;
				@fields = split (/\s+/, $line);
				while ($fields[0] eq "ATOM") {
					if ($fields[2] eq "CA") {
						print model_1
"$fields[5]\t$fields[3]\t$fields[6]\t$fields[7]\t$fields[8]\n";
					}
					$line = <inFile>;
					chomp $line;
```

```
                                        @fields = split (/\s+/, $line);
                                }
                }
                elsif ($fields[0] eq "MODEL" and $fields[1] == "11") {
                                $line = <inFile>;
                                $line = <inFile>;
                                chomp $line;
                                @fields = split (/\s+/, $line);
                                while ($fields[0] eq "ATOM") {
                                        if ($fields[2] eq "CA") {
                                                print model_11
"$fields[5]\t$fields[3]\t$fields[6]\t$fields[7]\t$fields[8]\n";
                                        }
                                        $line = <inFile>;
                                        chomp $line;
                                        @fields = split (/\s+/, $line);
                                }
                }
        }

        close model_1;
        close model_11;
        close inFile;

        open outFile, ">$file.tcl";
        print outFile <<'VMD1';
# copy and paste the content of this file into VMD tcl console

display resize 1000 1000
draw color red

proc vmd_draw_arrow {mol start end {res 28}} {
        set middle [vecadd $start [vecscale 0.9 [vecsub $end $start]]]
        graphics $mol cylinder $start $middle radius 0.15 resolution $res filled yes
        graphics $mol cone $middle $end radius 0.35 resolution $res }

proc vmd_draw_delete {args} {
   set i 0
   foreach list $args {
      if {$i == 0} {
         lassign $list mol
      incr i
      }
      if {$list == "all"} {
         graphics $mol delete all
            return
```

```
      }
      foreach id $list {
         graphics $mol delete $id
      }
   }
   return
}

VMD1
        open arrow_start, "$file.m1";
        open arrow_end, "$file.m11";
        while (defined($line_1 = <arrow_start>)) {
                chomp $line_1;
                @f_start = split (/\s+/, $line_1);
                $line_2 = <arrow_end>;
                chomp $line_2;
                @f_end = split (/\s+/, $line_2);
                print outFile "draw arrow {$f_start[2] $f_start[3] $f_start[4]} {$f_end[2]
$f_end[3] $f_end[4]}\n";
        }
        close arrow_start;
        close arrow_end;
        close outFile;

}
```

A.1.6 NMA_RMSD_process.pl

*This is a perl script used to process the .poc files (pocket information) and extract the pocket of

interest. Then save as a pdb file.*

```
open outFile, ">pnbCE_mode11_poc.pdb";
$match = "80";
@files = <*.poc>;
@fields = [];
foreach $file (@files) {
        open inFile, $file;
        $line = <inFile>;
        while (defined($line = <inFile>)) {
                chomp $line;
                @fields = split (/\s+/, $line);
                if ($fields[11] eq $match) {
                        print outFile "$line\n"; }
        }
```

```
        close inFile;
        close outFile;
}
```

A.2 PYTHON SCRIPTS

A.2.1 NMA_RMSD_process.pl

This is a python script used to write a VMD script to draw arrows for NMA.

```
import string
script_writer = open('VMD_arrow_model_1to2.tcl', 'w')

file_start = open('model_1.bgf', 'r')
script_writer.write("draw color blue\n")
script_writer.write("proc vmd_draw_arrow {mol start end} {\n"
                "set middle [vecadd $start [vecscale 0.9 [vecsub $end $start]]]\n"
                "graphics $mol cylinder $start $middle radius 0.15\n"
                "graphics $mol cone $middle $end radius 0.35 }\n")

for line_end in open('model_2.bgf', 'r'):
                    list_2 = string.split(line_end)
                    line_start = file_start.readline()
                    list_1 = string.split(line_start)
            if list_1[0] == 'ATOM' and list_1[1] == list_2[1] :
                    x = "draw arrow {%f %f %f} {%f %f %f}\n" %(float(list_1[6]),
                                                               float(list_1[7]),
                                                               float(list_1[8]),
                                                               float(list_2[6]),
                                                               float(list_2[7]),
                                                               float(list_2[8]))
                    script_writer.write(x)
script_writer.close()
```

A.2.2 pnbCE_Gorge_Analysis.py

This is a python script used to monitor the diameter of the gorge during MD simulation.

```
import string
from math import *
from MMTK import *
from MMTK.PDB import PDBConfiguration
from Numeric import *
```

```
from MMTK.MolecularSurface import *
def set_atom_positions():
	offset = 0
	for atoms in universe.atomList():
		atoms.setPosition(coords[offset: offset+3])
		offset = offset + 3
		print atoms.position(None)

# First, read the PDB file with the initial coordinates
configuration = PDBConfiguration('test_2.pdb')
universe = InfiniteUniverse()
myprotein = configuration.createAll(None, 1)
universe.addObject(myprotein)
#universe.addObject(configuration.createAll(None, 1))
#print universe.objectList()

#Determine the number of atoms
no_atoms = len(universe.atomList())

# create a list with the gorge sizes to measure
gorgeList = []
for i in range(0,11):
	for j in range(0,10):
		gorgeList.append(0.1*float(i)+0.01*float(j))
gorgeList.sort()

#change the Van der Waal's radii to those used by SPOCK
vdw = {'H': 0.11, 'C':0.155,'N':0.14,'O':0.135,'S':0.181}
for atom in universe.atomList():
	atom.vdW_radius = vdw[atom.symbol]

# Open the Amber mdcrd file
#coords =[]
input = open('2QE3_2.md.x','r')
# Skip the first line: it's just a title
input.readline()

# set the number of frames to read
no_frames = 1
for j in range(no_frames):
	coords = []
	while len(coords) < no_atoms*3:
		inputline = string.split(input.readline())

		for i in range(len(inputline)):
			coords.append(string.atof(inputline[i])/10.0)
```

```
        # Skip the box line
        #input.readline()
        set_atom_positions()

        maxGorgeSize = 0.0
        for gorgeSize in gorgeList:
                total_area = 0.0
                #print "frame %s, size %s" %(i,gorgeSize)
                #pr = float(gorgeSize) + 0.0
                his399 = ser189 = 0
                protsurf = surfaceAtoms(myprotein, probe_radius = gorgeSize)
                for mykey in protsurf.keys():
                        key_string = "%s" % mykey
                        if string.find(key_string,'His399') >= 0:
                                his399 = 1
                        elif string.find(key_string,'Ser189') >= 0:
                                ser189 = 1
                                #print key_string + " , " + "%s" % protsurf[mykey]
                                #total_area = total_area + protsurf[mykey]
                mystring = "%s %s" % (ser189,his399)
                print "gorge: %s, residues found:" % gorgeSize,
                print mystring
                if ser189 and his399:
                        maxGorgeSize = gorgeSize
                else: break

                #print pr,total_area
        print "frame %s, max %s" % (j,maxGorgeSize)
#       outList.append("%s, %s\n" % (i+1, maxGorgeSize))
#outFile = open('race_gorge_long.dat','w')
#outFile.writelines(outList)
#outFile.close()
input.close()
```

A.2.3 GetAveCoords.py

*This is a python script used for making a cross-correlation matrix for the C-alpha carbons during

a dynamics run.*

```
import string
from math import *
from MMTK import *
from MMTK.PDB import PDBConfiguration, PDBOutputFile
```

```
from Numeric import *
#from MMTK.MolecularSurfaceNSC import *
from MMTK import MolecularSurface, MolecularSurfaceNSC
from MMTK.Proteins import Protein
from Scientific.Geometry import *

import os
#import math
import sys
#thishostnum = int(sys.argv[1])
#numHosts = 23.0
#numFrames = 10000.0
#frameSections = numFrames / numHosts
#myFramesStart = int((thishostnum-1)*frameSections + 0.5) +1
myFramesStart = 1
#myFramesEnd = myFramesStart + 200
myFramesEnd = 10
#int(thishostnum*frameSections + 0.5)

#print "node %i, frameStart %i, frameEnd %i" % ( thishostnum, myFramesStart, myFramesEnd )

def set_atom_positions(atoms_List):
        offset = 0
        for atoms in atoms_List :
                atoms.setPosition(Vector(coords[offset], coords[offset+1], coords[offset+2]))
                offset = offset + 3

# First, read the PDB file with the initial coordinates
#configuration = PDBConfiguration('2QE3_2_mdstart.pdb')
universe = InfiniteUniverse()
universe.protein = Protein('2QE3_2_mdstart.pdb')
#myprotein = configuration.createAll(None, 1)
myprotein = universe.protein
#universe.addObject(myprotein)
# this next statement is to force MMTK to assign indices to all atoms
universe.configuration()
#universe.protein = Protein(myprotein)

# create a list of the atoms in PDB order
natoms = universe.numberOfAtoms()
atoms = natoms*[None]
for atom in universe.atomList():
        atoms[atom.index] = atom
        #aindex = "%s" % atom.index
        #print atom, string.atoi(aindex)
        #print atom.name
```

```
# now, the list "atoms" will be in the same order as the PDB file.
# here's a list of the C_alpha from the PDB file
CA_List = []
for a in atoms:
        if a.name == 'C_alpha' :
                CA_List.append(a)
#Determine the number of atoms
no_atoms = len(universe.atomList())

# Open the Amber mdcrd file
input = open('2QE3_2.md.x','r')
# Skip the first line: it's just a title
input.readline()

#Skip down to the start of the frame you want

j = 1
while j < myFramesStart:
        coords = []
        while len(coords) < no_atoms*3:
                inputline = string.split(input.readline())
                for i in range(len(inputline)):
                        coords.append(string.atof(inputline[i])/10.0)
        j=j+1

# Set up the output
#outFile = open('/home/analysis/2QE3_7.md.newdat'+str(thishostnum),'w')
outFile = open('atom_CA_ave.pdb', 'w')
ave_coords = []
for j in range(3*len(CA_List)):
        ave_coords.append(0.0)
# Start reading frames
total_frames = float(myFramesEnd + 1 - myFramesStart)
for j in range(myFramesStart, myFramesEnd+1):
        coords = []
        while len(coords) < no_atoms*3:
                inputline = string.split(input.readline())

                for i in range(len(inputline)):
                        coords.append(string.atof(inputline[i])/10.0)
        # set the new atom positions from the coordinates
        set_atom_positions(atoms)
#       outFile2 = PDBOutputFile("test%s.pdb" % j)
#       outFile2.write(universe.protein)
#       outFile2.close()
```

```
        # get the CA coords and average
        index_coords = 0
        for CA_atoms in CA_List :
                xyz = CA_atoms.position()
                ave_coords[index_coords] = ave_coords[index_coords] + (xyz.x()/total_frames)
                ave_coords[index_coords+1] = ave_coords[index_coords+1] +
(xyz.y()/total_frames)
                ave_coords[index_coords+2] = ave_coords[index_coords+2] +
(xyz.z()/total_frames)
                index_coords = index_coords + 3

name = 'ATOM'
serial = 7524
residue = 'UNK'
res_no = 490
atom_type = 'CA'
index_coords = 0
for CA_atoms in CA_List :
        #xyz = CA_atoms.position()
        outFile.write( "%-6s%5s  %-4s%-4s %4d    %8.3f%8.3f%8.3f\n" %
(name,serial,atom_type,residue,res_no,ave_coords[index_coords]*10,ave_coords[index_coords+
1]*10,ave_coords[index_coords+2]*10) )
        index_coords = index_coords+3
        serial = serial + 1

#outString = "%s\t%s\n" % (j, maxGorgeSize)
#outFile.write( outString )
#       outFile.flush()
#       outList.append("%s, %s\n" % (i+1, maxGorgeSize))
#outFile.writelines(outList)
outFile.close()
input.close()
```

A.2.4 crosscorrel5.py

This is a python script used for constructing DCCM during a dynamics run.

```
import string
from math import *
from MMTK import *
from MMTK.PDB import PDBConfiguration, PDBOutputFile
from Numeric import *
from MMTK.Proteins import Protein
from Scientific.Geometry import *
```

```
import os
#import math
import sys

#thishostnum = int(sys.argv[1])
#numHosts = 23.0
#numFrames = 10000.0
#frameSections = numFrames / numHosts
#myFramesStart = int((thishostnum-1)*frameSections + 0.5) +1
myFramesStart = 1
#myFramesEnd = myFramesStart + 200
myFramesEnd = 10
#int(thishostnum*frameSections + 0.5)
#print "node %i, frameStart %i, frameEnd %i" % ( thishostnum, myFramesStart, myFramesEnd )

def set_atom_positions(atoms_List):
        offset = 0
        for atoms in atoms_List :
                atoms.setPosition(Vector(coords[offset], coords[offset+1], coords[offset+2]))
                offset = offset + 3

def matrix_reloaded(ma, fname) :
        # NOTE: the first line of every matrix file has to contain the number of frames used in
averaging
        matrix_file = open(fname,'r')
        frames = float(matrix_file.readline())
        if len(ma.shape) == 1 :
                for i in range(ma.shape[0]) :
                        ma[i] = float(matrix_file.readline())
        elif len(ma.shape) == 2 :
                for i in range(ma.shape[0]) :
                        splitline = string.split(matrix_file.readline())
                        for j in range(ma.shape[1]) :
                                ma[i,j] = float(splitline[j])
        else:
                matrix_problem = "The matrix %s is not supported" % fname
                raise matrix_problem
        return frames

def matrix_offloaded(ma, frames, fname) :
        matrix_file = open(fname,'w')
        matrix_file.write("%i\n" % frames)
        if len(ma.shape) == 1 :
                for i in range(ma.shape[0]) :
                        matrix_file.write("%f\n" % ma[i])
        elif len(ma.shape) == 2 :
```

```
                for i in range(ma.shape[0]) :
                        for j in range(ma.shape[1]) :
                                matrix_file.write("%f\t" % ma[i,j])
                        matrix_file.write("\n")
        else:
                matrix_problem = "The matrix %s is not supported" % fname
                raise matrix_problem

# Set flags for reading or writing
saveflag = 0
readflag = 0
for flag in sys.argv[1:] :
        if flag == '-s' : saveflag = 1
        elif flag == '-r' : readflag = 1
        else :
                raise "command line argument %s not recognized" % flag

# First, read the PDB file with the initial coordinates
#configuration = PDBConfiguration('2QE3_2_mdstart.pdb')
universe = InfiniteUniverse()
universe.protein = Protein('2QE3_2_mdstart.pdb')
#myprotein = configuration.createAll(None, 1)
myprotein = universe.protein
#universe.addObject(myprotein)
# this next statement is to force MMTK to assign indices to all atoms
universe.configuration()
#universe.protein = Protein(myprotein)

# create a list of the atoms in PDB order
natoms = universe.numberOfAtoms()
atoms = natoms*[None]
for atom in universe.atomList():
        atoms[atom.index] = atom

# now, the list "atoms" will be in the same order as the PDB file.
# here's a list of the C_alpha from the PDB file
CA_List = []
for a in atoms:
        if a.name == 'C_alpha' :
                CA_List.append(a)

# Get the average CA coordinates from an external file
CA_input = open('2QE3_atom_CA_ave.pdb','r')
CA_ave_coords_V = []
for i in range(len(CA_List)):
        CA_inputline = CA_input.readline()
```

```
        if not CA_inputline :
                CA_input_Error = "something is wrong with the CA file!"
                raise CA_input_Error
        CA_splitline = string.split(CA_inputline)
        r_ave = Vector(float(CA_splitline[5])/10.0, float(CA_splitline[6])/10.0,
float(CA_splitline[7])/10.0)
        CA_ave_coords_V.append(r_ave)
        # Note that this should be the same order as the CA's in the PDB file

#Determine the number of atoms
no_atoms = len(universe.atomList())

# set up an array to hold the dynamic cross-correlation data
# this uses the zeros() function from Numeric
DCC_array = zeros((len(CA_List), len(CA_List)), Float)
# and three arrays to hold the delta_r values for the CA_atoms
delta_r = zeros((len(CA_List)), Float)
delta_r2= zeros((len(CA_List)), Float)
dri_drj= zeros((len(CA_List), len(CA_List)), Float)

# What are the total number of frames to used?
total_frames = float(myFramesEnd + 1 - myFramesStart)

# See if you need to read in earlier matrices
if readflag :
        matrix_reloaded(dri_drj, 'dri_drj_matrix')
        #old_frames = matrix_reloaded(delta_r, 'delta_r_matrix')
        old_frames = matrix_reloaded(delta_r2, 'delta_r2_matrix')
        total_frames = old_frames + total_frames
        for i in range(len(CA_List)) :
                #delta_r2[i] = delta_r2[i] + (delta_r[i]*delta_r[i])/total_frames
                delta_r2[i] = delta_r2[i]*old_frames/total_frames
                for j in range(len(CA_List)) :
                        dri_drj[i,j] = dri_drj[i,j]*old_frames/total_frames

# Begin working with the amber coordinates file
# Open the Amber mdcrd file
input = open('2QE3_2.md.x','r')
# Skip the first line: it's just a title
input.readline()

#Skip down to the start of the frame you want

j = 1
while j < myFramesStart:
        coords = []
```

```
        while len(coords) < no_atoms*3:
                inputline = string.split(input.readline())
                for i in range(len(inputline)):
                        coords.append(string.atof(inputline[i])/10.0)
        j=j+1

# Start reading frames
#total_frames = float(myFramesEnd + 1 - myFramesStart)
for j in range(myFramesStart, myFramesEnd+1):
        coords = []
        while len(coords) < no_atoms*3:
                inputline = string.split(input.readline())

                for i in range(len(inputline)):
                        coords.append(string.atof(inputline[i])/10.0)

        # set the new atom positions from the coordinates

        set_atom_positions(atoms)

        # Calculate the delta_r and delta_r2 values for this frame
        for i_coords in range(len(CA_List)) :
                delta_r[i_coords] = (CA_List[i_coords].position() -
CA_ave_coords_V[i_coords]).length()
                delta_r2[i_coords] = delta_r2[i_coords] +
(delta_r[i_coords]*delta_r[i_coords])/total_frames

        # Calculate the dri_drj matrix
        for i_coords in range(len(CA_List)) :
                for j_coords in range(len(CA_List)) :
                        dri_drj[i_coords,j_coords] = dri_drj[i_coords,j_coords] +
(delta_r[i_coords]*delta_r[j_coords])/total_frames

# The loop over frames ends here
input.close()

# Now offload the matrices if needed.
#  Note that you have to offload before taking the sqrt as you want the sqrt of the final sum
if saveflag :
        matrix_offloaded(dri_drj, total_frames, 'dri_drj_matrix')
        #matrix_offloaded(delta_r, 'delta_r_matrix')
        matrix_offloaded(delta_r2, total_frames, 'delta_r2_matrix')

# Now take the square root of the delta_r2 matrix
for i in range(len(CA_List)) :
        delta_r2[i] = sqrt(delta_r2[i])
```

```
# calculate the DCC matrix and write it out
# Set up the output
DCC_outFile = open('2QE3_2.md.DCC', 'w')
for i in range(len(CA_List)) :
        for j in range(len(CA_List)) :
                DCC_array[i,j] = dri_drj[i,j]/(delta_r2[i]*delta_r2[j])
                DCC_outFile.write("%f\t" % DCC_array[i,j])
        DCC_outFile.write("\n")
DCC_outFile.close()
```

A.3 OTHER SCRIPTS

A.3.1 runmd_3.in

This is an AMBER script used for production simulation.

```
 run MD at 300K
&cntrl
        imin=0,
        ntc=2, ntf=2,
        cut=12.0, igb=1, saltcon=0.2, gbsa=1,
        ntpr=50, ntwx=50, ntwe=50,
        nstlim=500000, dt=0.002,
        ntt=1, tempi=300.0, temp0=300.0, tautp=2.0,
        ntx=5, irest=1, ntb=0, ntwr=500,
      nscm=500,
&end
```

A.3.2 pnbCE_RMSD_10ns.ptraj

*This is an AMBERtool script used for analyzing the output trajectory files and generating the

RMSD of the overall structure.*

```
trajin 2QE3_12.md.am8.x
trajin 2QE3_13.md.am8.x
trajin 2QE3_14.md.am8.x
trajin 2QE3_15.md.am8.x
trajin 2QE3_16.md.am8.x
trajin 2QE3_17.md.am8.x
trajin 2QE3_18.md.am8.x
trajin 2QE3_19.md.am8.x
```

```
trajin 2QE3_20.md.am8.x
trajin 2QE3_21.md.am8.x
reference 2QE3_2_mdstart.pdb
rms reference out pnbCE_RMSD_10ns.rms @CA time 0.1
```

A.3.3 pnbCE_RMSD_residue.ptraj

*This is an AMBERtool script used for analyzing the output trajectory files and generating the

RMSD of each residue in the protein structure.*

```
trajin 2QE3_12.md.am8.x
trajin 2QE3_13.md.am8.x
trajin 2QE3_14.md.am8.x
trajin 2QE3_15.md.am8.x
trajin 2QE3_16.md.am8.x
trajin 2QE3_17.md.am8.x
trajin 2QE3_18.md.am8.x
trajin 2QE3_19.md.am8.x
trajin 2QE3_20.md.am8.x
trajin 2QE3_21.md.am8.x
reference 2QE3_2_mdstart.pdb
rms reference out pnbCE_RMSD_12.rms @CA time 0.1
atomicfluct out pnbCE_RMSD_residue.rms @CA byres
atomicfluct out pnbCE_RMSD_residue.bfactor @CA byres bfactor
```

A.3.4 active_site_fluc.ptraj

*This is an AMBERtool script used for analyzing the output trajectory files and generating the

RMSD of the active site residues during 10 ns simulation.*

```
trajin 2QE3_12.md.am8.x
trajin 2QE3_13.md.am8.x
trajin 2QE3_14.md.am8.x
trajin 2QE3_15.md.am8.x
trajin 2QE3_16.md.am8.x
trajin 2QE3_17.md.am8.x
trajin 2QE3_18.md.am8.x
trajin 2QE3_19.md.am8.x
trajin 2QE3_20.md.am8.x
trajin 2QE3_21.md.am8.x
reference 2QE3_2_mdstart.pdb
rms reference out pnbCE_362.rms time 0.1 :362
```

```
rms reference out pnbCE_189.rms time 0.1 :189
rms reference out pnbCE_218.rms time 0.1 :218
rms reference out pnbCE_219.rms time 0.1 :219
rms reference out pnbCE_270.rms time 0.1 :270
rms reference out pnbCE_310.rms time 0.1 :310
rms reference out pnbCE_399.rms time 0.1 :399
```

A.3.5 arrow_mode7.tcl

This is a VMD tcl script used to generate the arrow plot for NMA.

```
display resize 1000 1000
draw color red
proc vmd_draw_arrow {mol start end {res 28}} {
        set middle [vecadd $start [vecscale 0.9 [vecsub $end $start]]]
        graphics $mol cylinder $start $middle radius 0.15 resolution $res filled yes
        graphics $mol cone $middle $end radius 0.35 resolution $res }

# vmd extension procedure
#
# $Id: vmd_draw_delete.tcl,v 1.1 2005/01/03 14:06:39 akohlmey Exp $
# replace the default 'draw delete' with a multibple argument
# (which may even be lists) aware version
proc vmd_draw_delete {args} {
   set i 0
   foreach list $args {
      if {$i == 0} {
           lassign $list mol
         incr i
      }
      if {$list == "all"} {
         graphics $mol delete all
            return
      }
      foreach id $list {
         graphics $mol delete $id
      }
   }
   return
}
```

A.4 PROTEIN SEQUENCES

```
       1  MTHQIVTTQY GKVKGTTENG VHKWKGIPYA KPPVGQWRFK APEPPEVWED
[coil_5]  MTHQIVTTQY GKVKGTTENG VHKWKGIPYA KPPVGQWRFK APEPPEVWED
[coil_21] MTHQIVTTQY GKVKGTTENG VHKWKGIPYA KPPVGQWRFK APEPPEVWED

      51  VLDATAYGPI CPQPSDLLSL SYTELPRQSE DCLYVNVFAP DTPSQNLPVM
[coil_5]  VLDATAYGPI CPQPS           RQSE DCLYVNVFAP DTPSQNLPVM
[coil_21] VLDATAYGPI CPQPSDLLSL SYTELPRQSE DCLYVNVFAP DTPSQNLPVM

     101  VWIHGGAFYL GAGSEPLYDG SKLAAQGEVI VVTLNYRLGP FGFLHLSSFD
[coil_5]  VWIHGGAFYL GAGSEPLYDG SKLAAQGEVI VVTLNYRLGP FGFLHLSSFD
[coil_21] VWIHGGAFYL GAGSEPLYDG SKLAAQGEVI VVTLNYRLGP FGFLHLSSFD

     151  EAYSDNLGLL DQAAALKWVR ENISAFGGDP DNVTVFGESA GGMSIAALLA
[coil_5]  EAYSDNLGLL DQAAALKWVR ENISAFGGDP DNVTVFGESA GGMSIAALLA
[coil_21] EAYSDNLGLL DQAAALKWVR ENISAFGGDP DNVTVFGESA GGMSIAALLA

     201  MPAAKGLFQK AIMESGASRT MTKEQAASTA AAFLQVLGIN ESQLDRLHTV
[coil_5]  MPAAKGLFQK AIMESGASRT MTKEQAASTA AAFLQVLGIN ESQLDRLHTV
[coil_21] MPAAKGLFQK AIMESGASRT MTKEQAASTA AAFLQVLGIN ESQLDRLHTV

     251  AAEDLLKAAD QLRIAEKENI FQLFFQPALD PKTLPEEPEK SIAEGAASGI
[coil_5]  AAEDLLKAAD QLRIAEKENI FQLFFQPALD PKTLPEEPEK SIAEGAASGI
[coil_21] AAEDLLKAAD QLRIAEKENI FQLFFQPALD PKTLPEEPEK SIAEGAASGI

     301  PLLIGTTRDE GYLFFTPDSD VHSQETLDAA LEYLLGKPLA EKAADLYPRS
[coil_5]  PLLIGTTRDE GYLFFTPDSD VHSQETLDAA LEYLLGKPLA EKAADLYPRS
[coil_21] PLLIGTTRDE GYLFFTPDSD VHSQETLDAA LEYLLGKPLA EKAADLYPRS

     351  LESQIHMMTD LLFWRPAVAY ASAQSHYAPV WMYRFDWHPE KPPYNKAFHA
[coil_5]  LESQIHMMTD LLFWRPAVAY ASAQSHYAPV WMYRFDWHPE KPPYNKAFHA
[coil_21] LESQIHMMTD LLFWRPAVAY ASAQSHYAPV WMYRFDWHPE KPPYNKAFHA

     401  LELPFVFGNL DGLERMAKAE ITDEVKQLSH TIQSAWITFA KTGNPSTEAV
[coil_5]  LELPFVFGNL DGLERMAKAE ITDEVKQLSH TIQSAWITFA KTGNPSTEAV
[coil_21] LELPFVFGNL DGL        ITDEVKQLSH TIQSAWITFA KTGNPSTEAV

     451  NWPAYHEETR ETVILDSEIT IENDPESEKR QKLFPSKGE*
[coil_5]  NWPAYHEETR ETVILDSEIT IENDPESEKR QKLFPSKGE*
[coil_21] NWPAYHEETR ETVILDSEIT IENDPESEKR QKLFPSKGE*
```

A.5 DNA SEQUENCES

```
       1  ATGACTCATC AAATAGTAAC GACTCAATAC GGCAAAGTAA AAGGCACAAC
[coil_5]  ATGACTCATC AAATAGTAAC GACTCAATAC GGCAAAGTAA AAGGCACAAC
[coil_21] ATGACTCATC AAATAGTAAC GACTCAATAC GGCAAAGTAA AAGGCACAAC

      51  GGAAAACGGC GTACATAAGT GGAAAGGCAT CCCCTATGCC AAGCCGCCTG
[coil_5]  GGAAAACGGC GTACATAAGT GGAAAGGCAT CCCCTATGCC AAGCCGCCTG
[coil_21] GGAAAACGGC GTACATAAGT GGAAAGGCAT CCCCTATGCC AAGCCGCCTG

     101  TCGGACAATG GCGTTTTAAA GCACCTGAGC CGCCTGAAGT GTGGGAAGAT
[coil_5]  TCGGACAATG GCGTTTTAAA GCACCTGAGC CGCCTGAAGT GTGGGAAGAT
[coil_21] TCGGACAATG GCGTTTTAAA GCACCTGAGC CGCCTGAAGT GTGGGAAGAT

     151  GTGCTTGATG CCACAGCGTA CGGCTCTATT TGCCCGCAGC CGTCTGATTT
[coil_5]  GTGCTTGATG CCACAGCGTA CGGCTCTATT TGCCCGCAGC CGTCT
[coil_21] GTGCTTGATG CCACAGCGTA CGGCTCTATT TGCCCGCAGC CGTCTGATTT

     201  GCTGTCACTT TCGTATACTG AGCTGCCCCG CCAGTCCGAG GATTGCTTGT
[coil_5]                              CG CCAGTCCGAG GATTGCTTGT
[coil_21] GCTGTCACTT TCGTATACTG AGCTGCCCCG CCAGTCCGAG GATTGCTTGT

     251  ATGTCAATGT ATTTGCGCCT GACACCCCAA GTAAAAATCT TCCTGTCATG
[coil_5]  ATGTCAATGT ATTTGCGCCT GACACCCCAA GTAAAAATCT TCCTGTCATG
[coil_21] ATGTCAATGT ATTTGCGCCT GACACCCCAA GTAAAAATCT TCCTGTCATG

     301  GTGTGGATTC ACGGAGGCGC TTTTTATCTA GGAGCGGGCA GTGAGCCATT
[coil_5]  GTGTGGATTC ACGGAGGCGC TTTTTATCTA GGAGCGGGCA GTGAGCCATT
[coil_21] GTGTGGATTC ACGGAGGCGC TTTTTATCTA GGAGCGGGCA GTGAGCCATT

     351  GTATGACGGA TCAAAACTTG CGGCACAGGG AGAAGTCATT GTCGTTACAT
[coil_5]  GTATGACGGA TCAAAACTTG CGGCACAGGG AGAAGTCATT GTCGTTACAT
[coil_21] GTATGACGGA TCAAAACTTG CGGCACAGGG AGAAGTCATT GTCGTTACAT

     401  TGAACTATCG GCTGGGGCCG TTTGGCTTTT TGCACTTGTC TTCATTTAAT
[coil_5]  TGAACTATCG GCTGGGGCCG TTTGGCTTTT TGCACTTGTC TTCATTTAAT
[coil_21] TGAACTATCG GCTGGGGCCG TTTGGCTTTT TGCACTTGTC TTCATTTAAT

     451  GAGGCGTATT CTGATAACCT TGGGCTTTTA GACCAAGCCG CCGCGCTGAA
[coil_5]  GAGGCGTATT CTGATAACCT TGGGCTTTTA GACCAAGCCG CCGCGCTGAA
[coil_21] GAGGCGTATT CTGATAACCT TGGGCTTTTA GACCAAGCCG CCGCGCTGAA

     501  ATGGGTGCGA GAGAATATTT CAGCGTTTGG CGGTGATCCC GATAACGTAA
[coil_5]  ATGGGTGCGA GAGAATATTT CAGCGTTTGG CGGTGATCCC GATAACGTAA
[coil_21] ATGGGTGCGA GAGAATATTT CAGCGTTTGG CGGTGATCCC GATAACGTAA

     551  CAGTATTTGG AGAATCCGCC GGCGGGATGA GCATTGCCGC GCTGCTTGCT
```

```
[coil_5]  CAGTATNNGG AGAATCCGCC GGCGGGATGA GCATTGCCGC GCTGCTTGCT
[coil_21] CAGTATTTGG AGAATCCGCC GGCGGGATGA GCATTGCCGC GCTGCTTGCT

     601  ATGCCTGCGG CAAAAGGCCT GTTCCAGAAA GCAATCATGG AAAGCGGCGC
[coil_5]  ATGCCTGCGG CAAAAGGNCT GTTCCAGAAA GCAATCATGG NAAGCGGCGC
[coil_21] ATGCCTGCGG CAAAAGGCCT GTTCCAGAAA GCAATCATGG AAAGCGGCGC

     651  TTCTCGAACG ATGACGAAAG AACAAGCGGC GAGCACCTCG GCAGCCTTTT
[coil_5]  TTCTCGAACG ATGACGAAAG AACAAGCGGC GAGCACCTCG GCAGNCTTTT
[coil_21] TTCTCGAACG ATGACGAAAG AACAAGCGGC GAGCACCTCG GCAGCCTTTT

     701  TACAGGTCCT TGGGATTAAC GAGGGCCAAC TGGATAAATT GCATACGGTT
[coil_5]  TACAGGTCCT TGGGATTAAC GAGGGCCAAC TGGATAAATT GCATACGGTT
[coil_21] TACAGGTCCT TGGGATTAAC GAGGGCCAAC TGGATAAATT GCATACGGTT

     751  TCTGCGGAAG ATTTGCTAAA AGCGGCTGAT CAGCTTCGGA TTGCAGAAAA
[coil_5]  TCTGCGGAAG ATTTGCTAAA AGCGGCTGAT CAGCTTCGGA TTGCAGAAAA
[coil_21] TCTGCGGAAG ATTTGCTAAA AGCGGCTGAT CAGCTTCGGA TTGCAGAAAA

     801  AGAAAATATC TTTCAGCTGT TCTTCCAGCC CGCCCTTGAT CCGAAAACGC
[coil_5]  AGAAAATATC TTTCAGCTGT TCTTCCAGCC CGCCCTTGAT CCGAAAACGC
[coil_21] AGAAAATATC TTTCAGCTGT TCTTCCAGCC CGCCCTTGAT CCGAAAACGC

     851  TGCCTGAAGA ACCAGAAAAA GCGATCGCAG AAGGGGCTGC TTCCGGTATT
[coil_5]  TGCCTGAAGA ACCAGAAAAA GCGATCGCAG AAGGGGCTGC TTCCGGTATT
[coil_21] TGCCTGAAGA ACCAGAAAAA GCGATCGCAG AAGGGGCTGC TTCCGGTATT

     901  CCGCTATTAA TTGGAACAAC CCGTGATGAA GGATATTTAT TTTTCACCCC
[coil_5]  CCGCTATTAA TTGGAACAAC CCGTGATGAA GGATATTTAT TTTTCACCCC
[coil_21] CCGCTATTAA TTGGAACAAC CCGTGATGAA GGATATTTAT TTTTCACCCC

     951  GGATTCAGAC GTTCATTCTC AGGAAACGCT TGATGCAGCG CTCGAGTATT
[coil_5]  GGATTCAGAC GTTCATTCTC AGGAAACGCT TGATGCAGCG CTCGAGTATT
[coil_21] GGATTCAGAC GTTCATTCTC AGGAAACGCN TGATGCAGCG CTCGAGTATT

    1001  TACTAGGGAA GCCGCTGGCA GAGAAAGTTG CCGATTTGTA TCCGCGTTCT
[coil_5]  TACTAGGGAA GCCGCTGGCA GAGAAAGTTG CCGATTTGTA TCCGCGTTCT
[coil_21] TACTAGGGAA GCCGCTGGCA GAGAAAGTTG CCGATTTGTA TCCGCGTTCT

    1051  CTGGAAAGCC AAATTCATAT GATGACTGAT TTATTATTTT GGCGCCCTGC
[coil_5]  CTGGAAAGCC AAATTCATAT GATGACTGAT TTATTATTTT GGCGCCCTGC
[coil_21] CTGGAAAGCC AAATTCATAT GATGACTGAT TTATTATTTT GGCGCCCTGC

    1101  CGTCGCCTAT GCATCCGCAC AGTCTCATTA CGCCCCTGTC TGGATGTACA
[coil_5]  CGTCGCCTAT GCATCCGCAC AGTCTCATTA CGCCCCTGTC TGGATGTACA
[coil_21] CGTCGCCTAT GCATCCGCAC AGTCTCATTA CGCCCCTGTC TGGATGTACA
```

```
    1151  GGTTCGATTG GCACCCGAAG AAGCCGCCGT ACAATAAAGC GTTTCACGCA
[coil_5]  GGTTCGATTG GCACCCGAAG AAGCCGCCGT ACAATAAAGC GTTTCACGCA
[coil_21] GGTTCGATTG GCACCCGAAG AAGCCGCCGT ACAATAAAGC GTTTCACGCA

    1201  TTGGAGCTTC CTTTTGTCTT TGGAAATCTG GACGGATTGG AACGAATGGC
[coil_5]  TTAGAGCTTC CTTTTGTCTT TGGAAATCTG GACGGATTGG AACGAATGGC
[coil_21] TTAGAGCTTC CTTTTGTCTT TGGAAATCTG GACGGATTG

    1251  AAAAGCGGAG ATTACGGATG AGGTGAAACA GCTTTCTCAC ACGATACAAT
[coil_5]  AAAAGCGGAG ATTACGGATG AGGTGAAACA GCTTTCTCAC ACGATACAAT
[coil_21]            ATTACGGATG AGGTGAAACA GCTTTCTCAC ACGATACAAT

    1301  CAGCGTGGAT CACGTTCGCC AAAACAGGAA ACCCAAGCAC CGAAGCTGTG
[coil_5]  CAGCGTGGAT CACGTTCGCC AAAACAGGAA ACCCAAGCAC CGAAGCTGTG
[coil_21] CAGCGTGGAT CACGTTCGCC AAAACAGGAA ACCCAAGCAC CGAAGCTGTG

    1351  AATTGGCCTG CGTATCATGA AGAAACGAGA GAGACGCTGA TTTTAGACTC
[coil_5]  AATTGGCCTG CGTATCATGA AGAAACGAGA GAGACGCTGA TTTTAGACTC
[coil_21] AATTGGCCTG CGTATCATGA AGAAACGAGA GAGACGCTGA TTTTAGACTC

    1401  AGAGATTACG ATCGAAAACG ATCCCGAATC TGAAAAAAGG CAGAAGCTAT
[coil_5]  AGAGATTACG ATCGAAAACG ATCCCGAATC TGAAAAAAGG CAGAAGCTAT
[coil_21] AGAGATTACG ATCGAAAACG ATCCCGAATC TGAAAAAAGG CAGAAGCTAT

    1451  TCCCTTCAAA AGGAGAATAA
[coil_5]  TCCCTTCAAA AGGAGAATAA
[coil_21] TCCCNTCAAN AGGAGAATAA
```